

Paikkatietoaineiston tarjoaminen linkitettynä tietona kohdepalvelua hyödyntäen

Eero Hietanen

Insinöörیتieteiden korkeakoulu

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi
diplomi-insinöörin tutkintoa varten Espoossa 9.11.2015.

Työn valvoja:

Professori Kirsi Virrantaus

Työn ohjaaja:

TkT Lassi Lehto



Aalto-yliopisto
Insinöörیتieteiden
korkeakoulu

Tekijä: Eero Hietanen		
Työn nimi: Paikkatietoaineiston tarjoaminen linkitettynä tietona kohdepalvelua hyödyntäen		
Päivämäärä: 9.11.2015	Kieli: Suomi	Sivumäärä: 6+62
Koulutusohjelma: Geomatiikka		
Professuuri: Geoinformatiikka M3002		
Työn valvoja: Professori Kirsi Virrantaus		
Työn ohjaaja: TkT Lassi Lehto		
<p>Paikkatieto ja siihen liittyvät paikkatietoinfrastruktuurit, jotka tukevat paikkatietoaineiston saavutettavuus ja käytettävyyttä, ovat tärkeä osa sekä paikallista, alueellista että globaalia päätöksentekoa. Laadukas paikkatieto on sekä kallista tuottaa että tärkeää, joten on kannattavaa käyttää resursseja olemassa olevan ja koko ajan lisääntyvän paikkatiedon saavutettavuuden ja käytettävyyden parantamiseksi.</p> <p>Tässä työssä on tutkittu mahdollisuutta tarjota paikkatietoaineiston kohteet linkitettynä tietona olemassa olevaa WFS-kohdepalvelua hyödyntäen. Testiaineistona tutkimuksessa toimii Maanmittauslaitoksen Paikannimirekisteri. Aihetta tutkittiin luomalla Paikannimet linkitettynä tietona -verkkopalvelu, joka tarjoaa Paikannimirekisterin aineiston linkitettynä tietona paikkatietokohteiden URI-tunnusten perusteella. Verkkopalvelu hyödyntää MML:n olemassa olevaa WFS-kohdepalvelua ja muuntaa kohdepalvelusta haetun tiedon reaaliaikaisesti linkitetyksi tiedoksi. Paikkatietokohteille on työssä annettu yksilöivät URI-tunnukset JHS 193 ”Paikkatiedon yksilöivät tunnukset” -suosituksen mukaisesti. Tutkimuksessa on selvitetty lisäksi, kuinka paikkatietoaineiston tietomalli voidaan ontologisoida hyödyntäen sekä paikkatiedolle tarkoitettuja että muita webin yleisesti käytössä olevia ontologioita.</p> <p>Tutkimuksessa toteutetulla ratkaisulla on paikkatietoaineisto mahdollista tarjota linkitettynä tietona. Paikkatietoaineiston tarjoaminen linkitettynä tietona parantaa aineiston saavutettavuutta ja käytettävyyttä. Paikkatietoaineisto on saavutettavissa yksinkertaisesti URI-tunnusten perusteella ja yhdistettävissä muuhun webin linkitettyyn tietoon. Aineisto on näin myös käytettävissä webin linkitetyn tiedon käsittelyyn tarkoitettujen työkalujen ja sovellusten avulla.</p>		
Avainsanat: paikkatieto, linkitetty tieto, WFS-kohdepalvelu, URI-tunnus, ontologia		

Author: Eero Hietanen

Title: Providing Geographic Dataset Available via Web Feature Service as
Linked Data

Date: 9.11.2015

Language: Finnish

Number of pages: 6+62

Degree programme: Geomatics

Professorship: Geoinformatics M3002

Supervisor: Professor Kirsi Virrantaus

Advisor: D.Sc. (Tech.) Lassi Lehto

Spatial datasets and related spatial data infrastructures are vital in a reasonable decision making process at the global, areal and local scale. Collecting and maintaining high quality geographic information is expensive. Thus, all actions to improve the accessibility and the usability of the existing geographic information should be taken.

In this study, an option to provide a spatial dataset as linked data is investigated. The aim was to use an existing Web Feature Service as the source of the data. The test dataset in the study is Place Name Register (PNR) of National Land Survey Finland. The study was conducted by implementing a prototype web service, which provides the dataset as linked data. Making an ontology from the data model of the geographic dataset has also been studied. Uniform Resource Identifiers (URI) have been given to spatial objects according to the JHS 193 "Unique identifiers of the geographic information" recommendation. The idea of the web service is that all the data objects are available from the web via those URIs. The service performs an on-the-fly transformation to data from GML model to RDF model according to the created ontology.

The implementation solution used in the study enables delivering the content of a geographic dataset as linked data. Providing a dataset as linked data improves the usability and the accessibility of the geographic dataset. With this solution, the geographic dataset is available from the URIs of the data objects and can be connected to other linked data in the web. Data of the geographic dataset can also be processed using common linked data tools and applications of the web.

Keywords: geographic data, linked data, WFS, URI, ontology

Esipuhe

Tämä diplomityö on tehty Maanmittauslaitoksen Paikkatietokeskuksessa (FGI) osana maa- ja metsätalousministeriön, Suomen ympäristökeskuksen ja Maanmittauslaitoksen rahoittamaa HY-ID -projektia.

Haluaisin kiittää työni ohjaajaa TkT Lassi Lehtoa ja työni valvojaa Professori Kirsi Virrantausta. Kiitos myös Lassille ja Professori Tapani Sarjakoskelle mahdollisuudesta työskennellä opintojeni ohella ensin Geodeettisella laitoksella ja myöhemmin Paikkatietokeskuksessa. Lisäksi haluaisin kiittää työni aiheen parissa työskennelleitä Kai Koistista, Teemu Leskistä ja Esa Tiaista yhteistyöstä. Kiitokset myös Osma Suomiselle hyvistä vinkeistä työhöni liittyen sekä Christian Koskelle ruotsinkielen konsultaatiosta ja vertaisavusta. Iso kiitos tutkimusryhmämme jäsenille Pekka Latvalalle ja Jaakko Kähköselle arvokkaasta avusta kaikkiin työtehtäviin liittyen sekä myös muille Paikkatietokeskuksen työntekijöille hyvän työilmapiirin luomisesta.

Haluaisin kiittää myös kaikkia tärkeitä ihmisiä saamastani tuesta viimeisen neljän vuoden opiskelujakson aikana. Kiitos kaikille katrillilaisille fyysisen ja henkisen kunnon ylläpidosta. Kiitokset ystävilleni ja sukulaisilleni, ilman teitä en olisi tässä nyt. Kiitos isille, Pirjolle, Sannille, Arille, Ying-Chanille ja isoäidille tärkeästä tuesta. Kiitos äidilleni, joka on aina jaksanut kannustaa ja olla ylpeä pojistaan.

Viimeisin kiitos kummilapsilleni Antille, Usvalle, Carolinalle ja Filippalle. Teissä on tulevaisuus!

Masala, 23.10.2015

Eero Hietanen

Sisällysluettelo

Tiivistelmä	ii
Tiivistelmä (englanniksi)	iii
Esipuhe	iv
Sisällysluettelo	v
Lyhenteet	vi
1 Johdanto	1
2 Paikkatietoinfrastruktuuri	4
2.1 Kohdepalvelu	4
2.2 GML-merkintäkieli	5
3 Semanttisen webin teknologiat ja linkitetty tieto	7
3.1 Linkitetyn tiedon periaatteet	7
3.2 URI-tunnukset ja uudelleenohjaukset	8
3.3 URI-tunnusten antaminen	10
3.4 RDF ja RDF-tiedon sarjallistaminen	12
3.5 SPARQL- ja GeoSPARQL -kyselykielet	15
3.6 RDF-skeema, RDF-sanastot ja OWL 2 -ontologiakieli	17
3.7 GeoSPARQL-ontologia	20
4 Aiemmat tutkimukset	24
5 Paikannimet linkitettynä tietona -palvelu	26
5.1 Paikannimirekisteri	26
5.2 PNR-ontologian luominen	30
5.3 Palvelun toteutus	37
5.3.1 Palvelun alusta ja käytetyt työkalut	37
5.3.2 URI-tunnusten ja uudelleenohjausten määrittäminen	38
5.3.3 Palvelun toimintaperiaate	40
5.3.4 Ontologian tarjoaminen	40
5.3.5 Paikka- ja paikannimikohteiden tarjoaminen	42
5.3.6 Aineistolinkitysten toteuttaminen	46
5.4 Palvelun hyöty ja käyttötapauksia	49
5.5 Ratkaisun arviointi ja jatkokehitysajatuksia	51
6 Yhteenveto	54
Viitteet	57

Lyhenteet

Lyhenteet

GIS	Geographic Information System
GML	Geographic Markup Language
HTTP	Hypertext Transform Protocol
IRI	Internationalized Resource Identifier
MML	Maanmittauslaitos
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
PNR	Paikannimirekisteri
RDF	Resource Description Framework
RDFS	RDF-skeema (engl. RDF Schema)
SDI	Paikkatietoinfrastruktuuri (engl. Spatial Data Infrastructure)
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol And RDF Query Language
URI	Uniform Resource Identifier
UML	Unified Modeling Language
VoID	Vocabulary of Interlinked Datasets
W3C	World Wide Web Consortium
WFS	WFS-kohdepalvelu (engl. Web Feature Service)
WMS	Karttakuvapalvelu (engl. Web Map Service)
WWW	World Wide Web (lyh. web)
XML	Extensible Markup Language
YK	Yhdistyneet kansakunnat

Nimiavaruuslyhenteet

dcterms	http://purl.org/dc/terms/
ex	http://example.org/
foaf	http://xmlns.com/foaf/0.1/
geo	http://www.opengis.net/ont/geosparql#
owl	http://www.w3.org/2002/07/owl#
pnr	http://paikkatiedot.fi/def/1001010/pnr#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
schema	http://schema.org/
wd	http://www.wikidata.org/entity/
xsd	http://www.w3.org/2001/XMLSchema#

1 Johdanto

Paikkatieto ja siihen liittyvät paikkatietoinfrastruktuurit (lyh. SDI), jotka tukevat paikkatietoaineiston saavutettavuutta ja käytettävyyttä, ovat tärkeä osa sekä paikallista, alueellista että globaalia päätöksentekoa. Paikkatietoinfrastruktuurit luovat edellytyksiä paikkatietoaineistojen laajamittaiselle käytölle erilaisten metatieto, lataus-, haku- ja muunnospalveluiden avulla. Toimiva paikkatietoinfrastruktuuri on välttämätön osa julkisen hallinnon toimia, kuten lainsäädäntöä, rahoitusta, palveluita, suunnittelua ja rakentamista. Laajan paikkatietoinfrastruktuurin ympärille syntyy myös kaupallisia palveluita, jotka hyödyntävät julkisen hallinnon paikkatietoja tarjotakseen niiden yhteyteen rakennettuja kohdennettuja palveluita. Yleisesti paikkatietoaineistojen tarjoamiseen tietoverkossa käytetään Open Geospatial Consortium:n (lyh. OGC) standardoimaa Web Feature Service (lyh. WFS) -kohdepalvelua. WFS-kohdepalvelun etuna verrattuna esimerkiksi tiedostolatauspalveluihin on, että haettu tieto voidaan tuottaa suoraan aineistontarjoajan tietokannasta kohdekohteisesti. Tällöin tieto on aina ajantasaista ja kohdennettua. Luodaksemme lisää yhteiskuntaa hyödyttävää kaupallista toimintaa paikkatietoinfrastruktuurien ympärille, on aineistojen laatua ja saavutettavuutta kehitettävä vastaamaan muuttuvia tarpeita.

Aineistojen saavutettavuuteen ja käytettävyyteen vaikuttavat käytettävät teknologiat, tietomallit ja aineistoista tarjolla olevat metatiedot. Paikkatietoinfrastruktuurit ovat rakentuneet vahvasti omien standardien varaan perinteisten sovellusalojen, kuten maanmittauksen ja rakentamisen suunnittelun, ohjatessa niiden toimintaa. Paikkatiedon erityisyys verrattuna muuhun tietoon asettaa omat vaatimuksensa käytettävälle tietomalleille. Paikkatieto jää tällöin helposti eristyksiin omiin tietosiiloihin. Paikkatietoja täytyy kuitenkin voida integroida laajemmin muihin tietoaaineistoihin yleisesti käytössä olevien tekniikoiden ja sovellusten avulla, jotta erilaiset palveluntarjoajat voivat hyödyntää aineistoja entistä paremmin. Yksi tällaista integrointia tukevista toimista on tuoda paikkatiedot linkitettynä tietona osaksi semanttista webiä (Schade, Granell ym. 2010). Tätä varten on luotu paljon erilaisia aloitteita, joista viimeaikaisia ovat mm. OGC:n ja World Wide Web Consortium:n (lyh. W3C) yhteinen Spatial Data on the Web Working Group -työryhmä (W3C 2015b) ja OGC:n vuonna 2012 julkaisema GeoSPARQL-standardi (OGC 2012b). Linkitetty tieto mahdollistaa paikkatiedon hyödyntämisen paremmin mm. tiedon haussa. Esimerkiksi linkittämällä uutisartikkeleita olemassa olevaan paikkatietoon voimme hakea uutisia sijainnin perusteella. Erilaisten tapahtumien ja ilmiöiden linkittäminen paikkatietoon mahdollistaa myös niiden paremman ymmärtämisen ja tutkimisen. Käytännön tasolla paikkatietoaineistojen julkaisemisesta linkitettynä tietona ei kuitenkaan ole paljon kokemuksia. Tässä tutkimuksessa pyritään löytämään ja kokeilemaan niitä käytänteitä, joilla paikkatietoaineisto olisi mahdollista ja järkevä julkaista linkitettynä tietona.

Linkitetty tieto on tapa toteuttaa semanttista webiä käytännössä. Semanttisen webin ideana on kehittää Internetin WWW-palvelua (myöhemmin web) älykkäämmäksi kehittämällä teknologioita, joilla webin sisällölle annetaan koneluettavia merkityksiä.

Toiminta-ajatuksena on, että webin sisältöön liitetään tietoa asioiden välisistä suhteista. Tämä mahdollistaa tiedon tuottamisen (ainakin osittaisen) automatisoinnin. Semanttisen webin keskeisiä teknologioita ovat URI-tunnukset, RDF-tietomalliin pohjautuva linkitetty tieto, sekä OWL-ontologiakieli.

Tällä hetkellä suomalaisia paikkatietoaineistoja ei tarjota laajamittaisesti linkitettynä tietona. Suomalaisia paikkatietoaineistoja on aiemmin muunneltu RDF-tiedoksi (esim. Linked Data Finland 2015 ja Lindroos 2008) ja paikkatiedoille on luotu ontologioita tiedonhaun tehostamiseksi (kts. Koistinen 2011). RDF-tietovarastojen luomisessa ei Suomessa ole kuitenkaan hyödynnetty paikkatietojen WFS-kohdepalveluita ajantasaisen aineiston tarjoamiseksi ja esimerkiksi OGC:n suosittamaa GeoSPARQL-standardia ei ole käytetty.

SPARQL-kyselykieli ja aineistoja välittävät SPARQL-palvelut ovat tärkeitä osia semanttisen webin infrastruktuurissa. Tschirner ym. (2011) ovat luoneet palvelun, jolla voidaan hakea paikkatietokohteita linkitettynä tietona W3C:n määrittelemän SPARQL-kyselykielen avulla. Palvelun erikoisuutena on se, että se hakee SPARQL-kyselyyn tarvittavan paikkatiedon suoraan WFS-kohdepalvelusta. Tieto muunnetaan reaaliaikaisesti linkitetyn tiedon graafiksi, josta kyselylle löydetään toivottu hakutulos. Näin tarjottavaa paikkatietoaineiston tietoa voidaan yhdistellä rajattomasti kaikkien webin SPARQL-palveluiden tarjoamien tietojen kanssa.

SPARQL-palveluiden lisäksi tärkeässä osassa ovat myös resursseille (tässä tapauksessa paikkatietokohteille) määritetyt pysyvät URI-tunnukset. URI-tunnusten perusteella kohteet ovat saatavilla suoraan URIn määrittelemästä osoitteesta HTTP-protokollaa hyödyntäen (W3C 2008). Tässä työssä tutkitaan, voidaanko luoda verkkopalvelu, joka tarjoaa paikkatietoaineiston kohteet linkitettynä tietona paikkatietokohteille määritettyjen yksilöllisten URI-tunnusten perusteella reaaliaikaisesti WFS-kohdepalvelua hyödyntäen. Tutkimuksessa selvitetään lisäksi, kuinka paikkatietoaineiston tietomalli voidaan ontologisoida OWL-ontologiakiielellä hyödyntäen GeoSPARQL-ontologiaa ja muita yleisesti käytössä olevia ontologioita. Tässä yhteydessä tutkitaan myös mahdollisuutta luoda paikkatietokohteista yksinkertaistettu malli käyttäen Schema.org-sanastoa, jotta yleiset hakukoneet tai karttapalvelut voisivat hyödyntää aineistoa paremmin. Yhtenä tavoitteena on selvittää, ovatko paikkatietokohteet saavutettavissa palvelun myötä yleisten hakukoneiden, kuten Googlen, kautta.

Testiaineistona tutkimuksessa toimii Maanmittauslaitoksen (lyh. MML) Paikannimirekisteri (lyh. PNR). Tutkimuskysymyksiin vastataan luomalla prototyypipalvelu, joka tarjoaa Paikannimirekisterin aineiston linkitettynä tietona reaaliaikaisesti hyödyntäen MML:n WFS-kohdepalvelua. Paikkatietokohteille määritellään pysyvät URI-tunnukset uuden ”JHS 193 – Paikkatiedon yksilöivät tunnukset” (JUHTA 2015) -suosituksen mukaisesti. Prototyypipalvelua varten PNR:n tietomalli ontologisoidaan hyödyntäen OWL-ontologiakieltä. Paikkatietokohteet tarjotaan luodun PNR-ontologian mukaisessa muodossa URI-tunnusten osoittamissa verkko-osoitteissa HTTP-protokollan avulla. RDF-tiedon laajamittaisen hyödyntämisen parantamiseksi tutkitaan myös aineiston paikkatietokohteiden linkittymistä toisiinsa luomalla

linkkejä kohteiden välille. Paikkatietoaineiston saavutettavuuden ja käytettävyyteen vaikuttavat lisäksi aineistosta saatavilla olevat metatiedot. Paikkatietoaineiston metatiedot tarjotaan prototyyppipalvelussa linkitettynä tietona hyödyntäen tarkoitukseen sopivia webin yleisiä RDF-sanastoja.

Tämän työn rakenne koostuu teoriaosuudesta ja käytännön toteutuksen kuvauksesta. Teoriaosuuden aluksi luvussa 2 esitellään paikkatietoinfrastruktuurin käsite ja siihen liittyviä keskeisiä teknologioita, kuten WFS-kohdepalvelu. Teoriaosuus jatkuu semanttisen webin ja linkitetyn tiedon teknologioiden esittelyllä luvussa 3 ja aiheeseen liittyvien tieteellisten tutkimusten katsauksella luvussa 4. Luvussa 5 kuvataan käytännön toteutus, joka sisältää MML:n Paikannimirekisterin tietomallin esittelyn sekä kuvauksen tietomallin ontologisoinnista ja luodun verkkopalvelun toteutuksen. Samassa luvussa myös pohditaan erilaisia tapoja hyödyntää ja käyttää palvelua sekä arvioidaan toteutetun ratkaisun toimivuutta. Lopuksi luvussa 6 esitetään yhteenveto.

2 Paikkatietoinfrastruktuuri

Rio de Janeirossa vuonna 1992 pidetty YK:n ilmastokokous ja kokouksen myötä hyväksytty Agenda 21 (YK 1992) ovat olleet eräänlainen lähtölaukaus valtioiden omaksumalle kestävän kehityksen politiikalle. Paikkatiedon (engl. geographic information) merkitys sosiaalisten, ympäristöllisten ja taloudellisten ongelmien ratkaisemisessa ja päätöksenteon tukena nähtiin kokouksessa ratkaisevana. Vuonna 2002 Johannesburgissa YK:n maailmanlaajuisessa kokouksessa (engl. World Summit on Sustainable Development), joka tunnetaan myös nimellä Rio+10, onnistuttiin tuomaan esille digitaalisen paikkatiedon mahdollisuudet ja hyödyt kestävän kehityksen edistämiseksi käytännössä. Paikkatiedon tuottaminen on kuitenkin kallista, joten sopivaa tietoa ei aina ole saatavilla. Olemassa olevan ja koko ajan lisääntyvän paikkatiedon saavutettavuuden ja käytettävyyden parantamiseksi onkin perustettu lukemattomia kansallisia ja kansainvälisiä ohjelmia ja projekteja. (GSDI 2004, s. 6.)

Paikkatieto ja siihen liittyvät infrastruktuurit (eli paikkatietoinfrastruktuurit), jotka tukevat paikkatiedon saavutettavuutta ja käytettävyyttä, ovat tärkeitä sekä paikallisessa, alueellisessa että globaalissa päätöksenteossa. Paikkatietoinfrastruktuuri (engl. Spatial Data Infrastructure, lyh. SDI) käsittää paikkatietoaineiston, aineiston tarjoajan ja käyttäjät sekä aineistoon liittyvät erilaiset haku- ja latauspalvelut sekä sovellukset. SDI:ssä toimijoita ovat julkisen tai yksityisen sektorin osapuolet ja myös yksittäiset käyttäjät. (GSDI 2004, s. 8.) Tärkeimpiä paikkatietoinfrastruktuuria määritteleviä Suomea koskevia lakeja ovat Euroopan Unionin antama direktiivi Euroopan yhteisön paikkatietoinfrastruktuurin (INSPIRE) perustamisesta ja tämän kansallinen implementaatio, Laki ja Asetus paikkatietoinfrastruktuurista. INSPIRE-direktiivi on osa Euroopan ympäristöpolitiikkaa ja tähtää olemassaolevien aineistojen tarjoamiseen Euroopan laajuisesti yhtenäisessä muodossa keskenään yhteensopivien palvelurajapintojen avulla (EY 2007).

Paikkatietoinfrastruktuurin tärkeimpiä verkkopalvelustandardeja aineistojen jakamiseksi ovat OGC:n WFS-standardi (kts. OGC 2010b) ja Web Map Service (lyh. WMS) -standardi (kts. OGC 2006). WFS-kohdepalvelut ovat rajapintapalveluita, joiden avulla paikkatietokohteita voidaan hakea vektorimuodossa aineistontarjoajan tietovarastosta käyttäen spatiaalisia ja temaattisia hakuehtoja. WMS-palvelut puolestaan tarjoavat paikkatietoaineiston karttakuvana käyttäjän määrittelemien ehtojen mukaisesti. Seuraavissa aliluvuissa kerrotaan tarkemmin WFS-kohdepalvelusta (luku 2.1). Lisäksi perehdytään WFS-kohdepalvelussa käytettävään Geographic Markup Language (lyh. GML) -merkintäkieleen (luku 2.2). GML on tarkoitettu paikkatietoaineistojen välittämiseen ja tallentamiseen.

2.1 Kohdepalvelu

Tässä työssä verkkopalveluilla tarkoitetaan World Wide Web -pohjaisia rajapintapalveluita, jotka hyödyntävät Hypertext Transfer Protocol (lyh. HTTP) -protokollaa.

Protokolla perustuu asiakas-palvelin -arkkitehtuuriin. Ideana on, että asiakasohjelma avaa TCP-yhteyden palvelimelle ja lähettää HTTP-pyyntön (engl. HTTP request). Palvelin palauttaa pyynnönmukaisen vastauksen (engl. HTTP response). Kohdepalvelu-käsitteellä voidaan tarkoittaa myös tiedostolatauspalvelua, jolloin paikkatietokohteet ovat saatavilla tiedostona (esimerkiksi GML-muodossa). Tässä työssä kohdepalvelulla viitataan kuitenkin nimenomaan WFS-kohdepalveluun.

OGC:n standardoima WFS-kohdepalvelu on verkkopalvelu, joka tarjoaa standardimuotoisen pääsyn paikkatietoaineistoon. Pääsy on samanlainen riippumatta itse rajapinnan takana olevan tietovaraston toteutustavasta. Standardi määrittelee operaatiot palvelun metatietojen ja aineiston tietomallien hakemiseen sekä paikkatietokohteiden hakuun ja muokkaamiseen. Kyselyoperaatiot (engl. query operations) mahdollistavat paikkatietokohteiden tai niihin liittyvien ominaisuuksien arvojen hakemisen tietovarastosta. Kyselyoperaatiot tukevat OGC:n Filter Encoding -standardin (kts. OGC 2010a) mukaisia rajausehtoja, jotka käsittävät esim. loogisia ja spatiaalisia operaatioita. Transaction-operaatiot (engl. transaction operations) mahdollistavat tietoaaineistossa olevien kohteiden lisäämisen, muokkaamisen ja poistamisen WFS-kohdepalvelun kautta. (OGC 2010b, s. 1).

Tällainen palvelurajapinnan standardisointi mahdollistaa sovellukset, joiden avulla voidaan hyödyntää kaikkea sellaista paikkatietoa, joka on tarjolla WFS-kohdepalveluiden kautta. Tässä työssä paikkatietokohteiden muuntamisessa linkitetyksi tiedoksi hyödynnetään WFS-kohdepalvelun kysely-operaatiota GetFeature. GetFeature-kyselyn pakollisia parametreja ovat mm. palvelun versio ja haettavien paikkatietokohteiden kohdetyyppi. Vapaaehtoisia parametreja ovat esimerkiksi siirtomuoto (engl. output format), paikkatietokohteen tunnus, jota käytetään vain haettaessa yhtä tiettyä kohdetta, ja OGC:n Filter Encoding -standardin mukaiset rajausehdot. GetFeature-kysely lähetetään verkkopalveluun joko HTTP GET tai HTTP POST -pyyntönä. GET-pyyntössä parametrit välitetään merkittynä pyynnön URL-osoitteen perään avain-arvo-parina. POST-pyyntössä parametrit välitetään pyynnön sisältöosiossa Extensible Markup Language (lyh. XML) -merkintäkielellä. Myös GET-pyyntössä arvona käytetään XML-merkintäkieltä lähetettäessä Filter Encoding -standardin mukaisia parametreja.

2.2 GML-merkintäkieli

Aineistontarjoajan paikkatietoaineiston välittäminen vaatii rajapintapalvelun lisäksi käyttäjän ja tarjoajan välisen yhteisen sopimuksen siitä, missä muodossa aineisto toimitetaan. Aineistontarjoaja säilyttää vektorimuotoista paikkatietoaineistoa usein relaatiotietokannassa, josta se voidaan sarjallistaa monenlaiseen muotoon. Käytettyjä vektorimuotoista paikkatietoa tukevia sarjallistamismuotoja ovat Ecma Internationalin JSON-standardiin perustuva GeoJSON (Butler ym. 2008), XML-merkintäkieleen perustuva GML (OGC 2012a), ESRI Shapefile (ESRI 1998) ja OGC:n GeoPackage (OGC 2014). Näistä kaksi ensimmäistä ovat tekstimuotoisia ja siten myös

ihmisluettavia muotoja. Shapefile ja GeoPackage ovat sen sijaan binäärimuotoisia. WFS-kohdepalvelun oletussarjallistamismuoto on GML, mutta palvelu mahdollistaa myös muiden sarjallistamismuotojen käytön (OGC 2010b).

Extensible Markup Language -merkintäkieli (lyh. XML, kts. W3C 2006a) on tapa välittää ja tallentaa rakenteista tietoa. Se perustuu alunperin vuonna 1986 standardoituun SGML-merkintäkieleen (engl. Standard Generalized Markup Language). XML erittelee tiedon loogisen sisällön (eli semantiikan) ja fyysisen rakenteen. XML määrittelee tavan, jolla mitä tahansa tietoa voidaan merkitä nimettyjen elementtien avulla. XML-dokumentti koostuu elementeistä, jotka voivat olla sisäkkäisiä tai peräkkäisiä. Elementtien järjestys määrittää dokumentin rakenteen. Elementeille voidaan määrittellä myös attribuutteja ja elementit voivat sisältää viittauksia toisiin elementteihin. XML on metakieli, eli se ei määrittele käytettävissä olevia elementtejä. Dokumentin rakenne (esimerkiksi elementtien nimet) voidaan määrittellä XML-skeemojen (engl. XML Schema, kts. W3C 2012d) avulla. Skeemat määrittelevät XML-dokumentin rakenteen ja käytettävät tietotyypit. Skeemoissa voidaan myös kuvailla tarkemmin eri elementtien ja attribuuttien sisältöä. XML-dokumentista voi olla viittaus yhteen tai useampaan XML-skeematiedostoon, jolloin dokumenttia lukeva sovellus osaa tulkita dokumentin skeeman mukaisesti.

Geographic Markup Language on XML-pohjainen merkintäkieli, jolla voidaan sarjallistaa paikkatietokohteita. GML-merkintäkieli on standardoitu XML-kielioppi eli säännöstö. GML-muotoiset paikkatietokohteet voivat sisältää piste-, viiva- tai aluegeometrioita. Kielioppi on määritelty GML-skeemassa (engl. GML Schema). GML-dokumentti sisältää itse sarjallistettavat paikkatietokohteet. Käytännössä GML-skeemaa laajennetaan luomalla sovellusaloitteisia sovellusskeemoja (engl. application schema). Sovellusskeemoilla määritellään mitä geometrioita ja muita ominaisuuksia on tietyillä kohdetyypeillä, kuten esimerkiksi tiellä, järvellä ja maakunnalla. ISO 19136 -standardissa ja sitä vastaavassa OGC:n GML 3.2.1 -standardissa määritelty GML on yhteensopiva ISO 19109 standardissa (kts. ISO 2005) määritellyn General Feature Model -mallin kanssa. (OGC 2012a.) Laajasti käytettyjä yhteisesti sovittuja GML-sovellusskeemoja ovat esimerkiksi INSPIRE-direktiivin 34 eri teeman sisältämille kohdetyypeille määritetyt skeemat.

3 Semanttisen webin teknologiat ja linkitetty tieto

Semanttinen web perustuu ajatukselle, että dokumenttien linkittämisen sijaan linkitetään tietoa. Semanttisen webin ideana on luoda yhteinen kieli, jonka avulla webissä olevalle tiedolle annetaan merkityksiä ja yhteyksiä (W3C 2013b ja 2015a). Jos tieto voidaan jakaa yhteisen ja yleisen kielen avulla, niin se mahdollistaa myös tiedon loputtoman yhdistelemisen eri lähteistä.

Linkitetty tieto (engl. Linked Data) viittaa käytäntöihin, jolla semanttisen webin ideaa on lähdetty toteuttamaan käytännössä. Keskeisiä linkitetyn tiedon teknologioita ovat Resource Description Framework (lyh. RDF) -tietomalli, yksilöivät URI-tunnukset (engl. Uniform Resource Identifiers) ja HTTP-pyynnön otsakkeita hyödyntävä Content Negotiation -tekniikka, jonka avulla palvelinta voidaan ohjata palauttamaan tieto tietyssä formaatissa. Erittäin tärkeä osa semanttista webiä ja linkitettyä tietoa ovat myös sanastot ja ontologiat, joilla tietosisältöä voidaan määritellä. Web Ontology Language (lyh. OWL) -ontologiakieli määrittelee terminologian, jolla voidaan luoda ontologioita. Ontologiat ovat tapa jäsentää tietoa (esimerkiksi RDF-muodossa olevaa) ja luoda tiedolle merkityksiä. Seuraavissa aliluvuissa paneudutaan tarkemmin linkitetyn tiedon periaatteisiin ja linkitetyn tiedon eri teknologioihin.

3.1 Linkitetyn tiedon periaatteet

Linkitetyn tiedon periaatteilla viitataan yleensä Berners-Leen (2006) esittelemiін neljään sääntöön (Heath ym. 2011):

1. Käytä URI-tunnuksia asioiden niminä.
2. Käytä HTTP-muotoisia URI-tunnuksia.
3. Jos joku menee URIn osoitteeseen, tarjoa hyödyllistä tietoa hyödyntäen standardeja (RDF, SPARQL).
4. Sisällytä linkkejä toisiin URI-tunnuksiin, jotta ihmiset voivat löytää lisää tietoa.

Ensimmäisen kohdan termi ”asiat” voivat tarkoittaa web-dokumentin tai digitaalisen sisällön lisäksi mitä tahansa reaailimaailman kohdetta tai abstraktia asiaa (Heath ym. 2011). Näitä asioita kutsutaan web-arkkitehtuurissa yleisesti nimellä resurssi. URI-tunnusten käyttäminen reaailimaailman kohteiden ja abstraktien asioiden nimeämisessä voidaan ajatella ensimmäisenä askeleena kohti webin versiota 3.0, jossa webin sisältö nähdään toisiinsa linkittyneinä asioina ja tiedon osina. Web 1.0 on ollut pelkästään linkitettyjä dokumentteja, kuvia tai muita mediaformaatteja ja web 2.0 on tuonut myös palvelut laajemmin osaksi webiä.

Toisessa periaatteessa mainittu HTTP-protokolla on mekanismi, jonka avulla webin sisältö saadaan jaettua. Hyödynnettäessä URI-tunnuksissa nimenomaan HTTP-muotoa, toimii tunnus asialle myös yksikäsitteisenä osoitteena, jolla asiasta saadaan webin kautta tietoa. Tätä osoitetta voivat hyödyntää niin sovellukset kuin käyttäjätkin esimerkiksi web-selaimen avulla.

Tärkeää sovellusten ja käyttäjien kannalta on, että tieto tarjotaan URI-osoitteessa standardoidussa muodossa. Ihmisluettavaksi tarkoitettu selainohjelmien tulkitsema HTML-formaatti on mahdollistanut webin perinteisen sisällön tarjoamisen kaikille. Linkitetyn tiedon kolmas periaate kehottaa tarjoamaan rakenteisen tiedon RDF-muodossa. RDF on yksinkertainen graafimuotoinen tietomalli, joka on suunniteltu toimimaan verkkoympäristössä. Se on koneluettavaa ja mahdollistaa itse tiedon ja asioiden linkittämisen niistä kertovien dokumenttien linkittämisen sijaan. (Heath ym. 2011.)

Linkitetyn tiedon neljäs periaate on erittäin tärkeä. Yksittäisten tiedon osien linkittäminen toisiinsa ja sitä kautta koko maailmanlaajuiseen tietoavaruuteen on olennaista, jotta kaikki mahdollinen tieto tulee käyttäjien ja sovellusten saataville, eikä jää eristyksiin erillisiin tietosiiloihin. (Heath ym. 2011.)

Linkitetylle tiedolle ja linkitetylle avoimelle tiedolle on luotu luokittelujärjestelmä, jolla voidaan mitata sen käytettävyyttä. Berners-Lee (2006) on luonut viiden tähden järjestelmän, jossa tähtiä jaetaan tiedon ominaisuuksien perusteella. Yksi tähti annetaan, jos data on ylipäättänsä saatavilla webissä. Kaksi tähteä annetaan, jos se on rakenteisessa koneluettavassa muodossa (esimerkiksi Excel-muodossa tekstitiedoston sijaan). Kolme tähteä saa, jos tieto on avoimessa formaatissa (esimerkiksi CSV-muodossa Excel-muodon sijaan). Neljään tähteen vaaditaan, että käytettävät tunnukset perustuvat W3C:n standardeihin. Viides tähti annetaan, jos tieto linkittyy muihin aineistoihin. Hyvönen ym. (2014) ehdottavat järjestelmän laajentamista seitsemään tähteen. Kuudennen tähden saa, jos aineiston mukana on tarjolla aineistoa kuvaileva skeema ja seitsemännen tähden, jos aineisto on validoitu vastaamaan tätä skeemaa. Skeema helpottaa aineiston luonteen ymmärtämistä ja siten myös sen hyödyntämistä ja soveltamista eri tapauksissa.

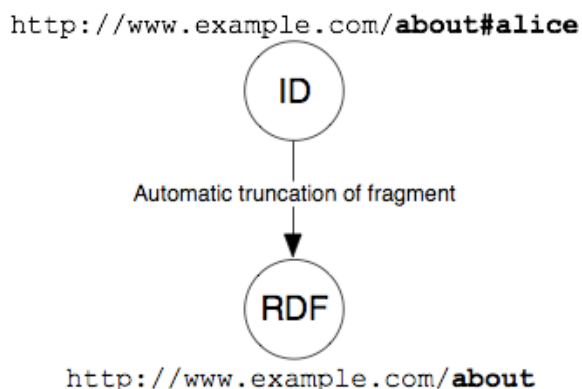
3.2 URI-tunnukset ja uudelleenohjaukset

Yleisesti URI-tunnukset muodostuvat peräkkäisistä hierarkkisesti järjestetyistä komponenteista. Skeema-osio (esimerkiksi "http") on pakollinen kaikille URI-tunnuksille. Skeema-osion jälkeen on "//"-alkuinen komponentti, joka määrittelee tunnuksen haltijan (engl. authority, esimerkiksi "paikkatiedot.fi"). Tätä osaa kutsutaan HTTP-muotoisissa URI-tunnuksissa nimellä verkkotunnus. Verkkotunnuksen jälkeen on polkuosio, joka koostuu "/"-merkillä erotetuista polkukomponenteista. Polkuosio on usein hierarkkisesti järjestetty ja se identifioi resurssin yhdessä skeeman ja verkkotunnuksen kanssa. Lisäksi URI-tunnuksessa voi olla kyselyosio (engl. query component), joka erotetaan "?"-merkillä ja joka sisältää usein avain-arvo-pareja. URI-tunnuksen

lopussa voi myös olla ns. ristikkomerkkiosio (engl. fragment), joka erotetaan ”#”-merkillä. Sekä kyselyosio että ristikkomerkkiosio ovat myös osa URI-tunnusta, vaikka niillä on lisäksi oma merkityksensä, kun tehdään HTTP-pyyntö URI-tunnuksen mukaiseen osoitteeseen asiakasohjelmalla. (Berners–Lee ym. 2005.)

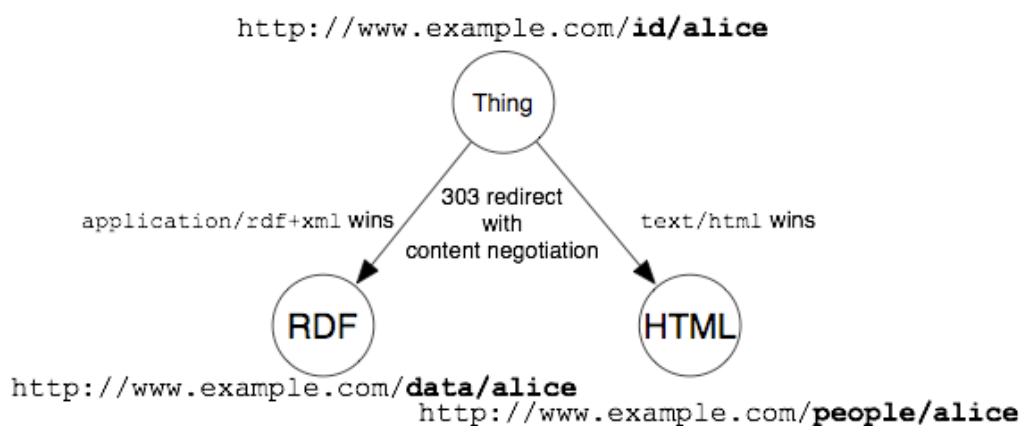
Annettaessa reaali maailman kohteelle tai abstraktille asialle URI-tunnus täytyy yleensä perustaa myös toinen URI-tunnus. Tämä toinen URI-tunnus on web-dokumentille, joka määrittelee ja kuvailee kyseisen kohteen. (W3C 2008). Erillisten URI-tunnusten käyttö mahdollistaa erillisen viittaamisen kohteeseen ja sitä määrittävään web-dokumenttiin. Tällainen toisen URI-tunnuksen tarvitsema resurssi voi olla esimerkiksi paikkatietokohde. Paikkatietokohteelle perustetaan siis lisäksi toinen URI-tunnus (kohteeseen liittyvä web-dokumentti), joka on esimerkiksi GML-muotoinen määrittely tai HTML-muotoinen kuvaus kohteesta. Jos HTML-muotoiseen kuvaukseen lisätään esimerkiksi kartta kohteen sijainnista, niin web-dokumentin viimeinen muokkausajankohta muuttuu. Itse paikkatietokohteen viimeinen muokkausajankohta ei kuitenkaan muutu. Kahdella erillisellä URI-tunnuksella voidaan ilmaista itse kohteeseen liittyviä tai kohdetta määrittelevään web-dokumenttiin liittyviä tietoja yksiselitteisesti ilman sekaannuksen vaaraa.

Resurssia määrittelevä tai kuvaileva web-dokumentti voidaan asettaa saataville resurssin URI-tunnuksen perusteella kahdella eri tavalla: hyödyntäen ristikkomerkki-URI-tunnusta (engl. Hash URI) tai uudelleenohjauksella (engl. Redirect, HTTP-vastauksen statuskoodi 303). Ristikkomerkki-URI-tunnuksen tapauksessa (kts. kuva 1) reaali maailman kohdetta tai abstraktia asiaa vastaava resurssi identifioidaan URI-tunnuksen ristikkomerkin jälkeisellä osiolla. Kun asiakasohjelmalla tehdään ristikkomerkki-URI-tunnuksen mukaiseen osoitteeseen HTTP-pyyntö, niin kyselyn vastaanottava palvelu hyödyntää ainoastaan ristikkomerkin etupuolella olevaa URI-tunnuksen osiota. URI-tunnuksen alkuosa palauttaa web-dokumentin, joka sisältää niiden kaikkien resurssien, joiden tunnukset ovat yhteneviä ristikkomerkin etupuolella olevan URI-tunnuksen osion kanssa. Tällöin haetun resurssin tietojen etsiminen palautetusta web-dokumentista jää asiakasohjelman tehtäväksi. (W3C 2008.)



Kuva 1: Ristikkomerkki-URI-tunnuksen hyödyntäminen. Lähde: W3C 2008.

Toinen tapa palauttaa resurssi on uudelleenohjaus. Uudelleenohjauksessa HTTP-pyyntö resurssin URI-tunnukseen palauttaa HTTP-vastauksen, joka sisältää pyynnön uudelleenohjauksen resurssia määrittelevän web-dokumentin URI-osoitteeseen. Koska resurssia määrittelevä web-dokumentti voi olla saatavilla useassa eri tietomuodossa, tarvitaan HTTP-protokollan mukaista Content Negotiation -tekniikkaa. Content Negotiation -tekniikka tarkoittaa HTTP-pyynnön otsakkeeseen sijoitetun Accept-attribuutin hyödyntämistä oikeanlaisen tietomuodon palauttamiseksi. Uudelleenohjaus voidaan tehdä Content Negotiation -tekniikkaa hyödyntäen sellaiseen URI-osoitteeseen, joka vastaa haluttua tietomuotoa (kts. kuva 2). On mahdollista käyttää myös yhtä yleistä URI-tunnusta kaikille web-dokumenteille, jotka määrittävät resurssia. Tällöin HTTP-pyyntö uudelleenohjataan tähän yleiseen URI-osoitteeseen, joka Content Negotiation -tekniikan mukaisesti palauttaa sisällön oikeassa muodossa (kts. kuva 3). Yleisen web-dokumentti-URI-tunnuksen tapauksessa palvelin myös lisää HTTP-vastauksen otsakkeeseen Content-Location-attribuutin, jossa määritellään oma URI-tunnus kyseiselle tietomuodolle. (W3C 2008.)

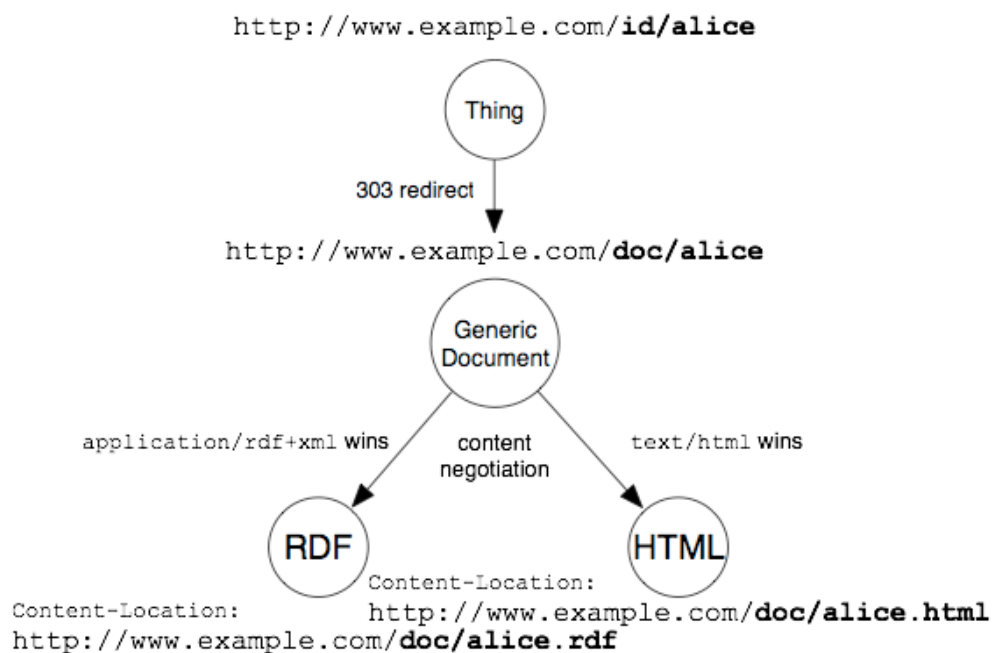


Kuva 2: Uudelleenohjaus Content Negotiation -tekniikan avulla. Lähde: W3C 2008.

URI-tunnusten sijasta linkitettyssä tiedossa käytetään myös URI-tunnuksen laajennosta Internationalized Resource Identifier (lyh. IRI) -tunnusta. Se mahdollistaa Unicode-merkistöstandardin hyödyntämisen tunnuksessa, jolloin URI-tunnukset eivät ole enää kieliriippuvaisia. Suomessa tämä laajennos näkyy esimerkiksi mahdollisuutena hyödyntää tunnuksissa skandinaavisia merkkejä. Tässä tutkimuksessa käytetään yleisesti termiä URI myös mahdollisista IRI-tunnuksista.

3.3 URI-tunnusten antaminen

Paikkatietokohteille yksilöllisten tunnusten luomiseksi on ollut monenlaisia käytäntöjä. Uusimmat INSPIRE-direktiiviin liittyvät tietotuotemäärittelyt ja niihin liittyvät tekniset puitedokumentit (esim. INSPIRE Drafting Team: Data Specifications 2013) velvoittavat yksilöllisten tunnusten julkaisemisen HTTP URI -muotoisena, sillä web nähdään tärkeimpänä paikkatiedon jakelukanavana. (Euroopan komissio 2015).



Kuva 3: Uudelleenohjaus yleiseen web-dokumentti-URI-osoitteeseen. Lähde: W3C 2008.

INSPIRE-direktiivin teknisessä General Conceptual Model -dokumentissa (INSPIRE Drafting Team: Data Specifications 2013) painotetaan reaali maailman kohteiden ja niitä vastaavien paikkatietokohteiden erottamista. Dokumentissa annetaan esimerkkinä neljä erilaista URI-tyyppiä, jotka voidaan tunnistaa niille määritellyistä polkukomponenteista `"/id"`, `"/so"`, `"/doc"` tai `"/def"`. `"/id"`-polkukomponentti viittaa siihen, että tunnus on reaali maailman kohteelle annettu URI-tunnus. Reaali maailmaa vastaavalle paikkatietokohteelle voidaan antaa `"/so"`-polkukomponentin sisältämä URI-tunnus. Dokumentti, joka määrittelee paikkatietokohdetta saa `"/doc"`-polkukomponentin tunnuksensa ja `"/def"`-polkukomponentti on tarkoitettu dokumenteille, jotka määrittelevät paikkatiedon käsitettä tai ontologiaa.

Julkisen hallinnon suositus JHS 193 Paikkatiedon yksilöivät tunnukset (JUHTA 2015) määrittelee INSPIRE-direktiivin vaatimuksiin perustuvan URI-tunnuksen rakenteen suomalaisille paikkatietokohteille. Paikkatietokohteille annettava `"/id"`-, `"/so"`- tai `"/def"`-polkukomponentin sisältävä URI-tunnus voi saada Maanmittauslaitoksen hallinnoiman `"paikkatiedot.fi"`-verkkotunnuksen. Tällaisella keskitetysti ylläpidettävällä verkkotunnuksella pyritään varmistamaan tunnuksen pysyvyys sekä jäljitettävyys. `"paikkatiedot.fi"`-verkkotunnus toimii uudelleenohjauspalveluna tiedontuottajan omiin palveluihin.

JHS 193 (JUHTA 2015) määrittelee paikkatietokohteen (`"/so"`) URI-tunnuksen rakenteen seuraavasti:

```

http://paikkatiedot.fi/so/{aineistotunnus}/{paikallinen_tunnus}
    [{versiotunnus}]
  
```

Paikkatietokäsitteen tunnus ("/def") määritellään:

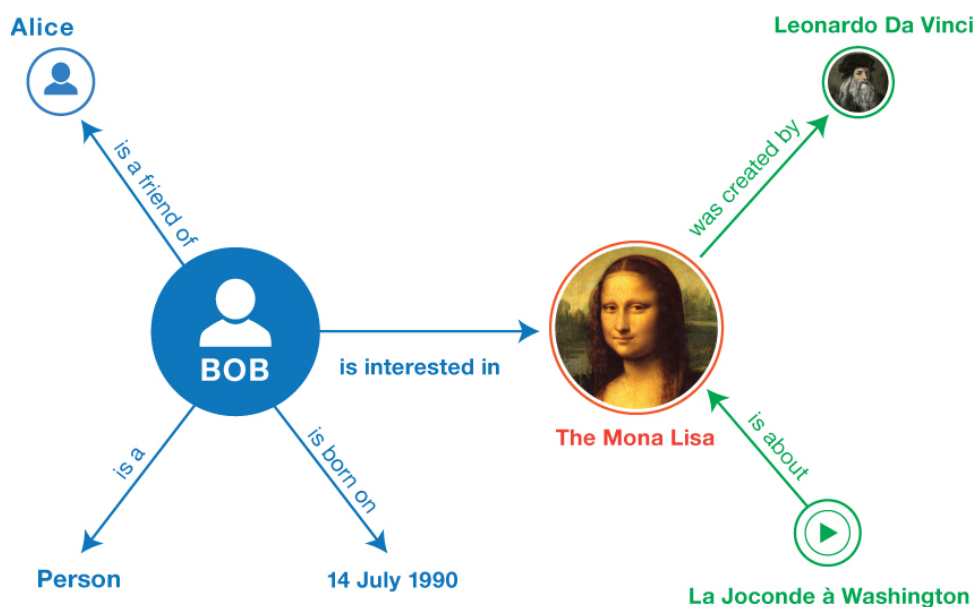
```
http://paikkatiedot.fi/def/{sanasto/skeema}/{paikallinen tunnus}
```

Paikkatietokohdetta tai käsitettä määrittelevän web-dokumentin URI-tunnus ("/doc") saa tiedontuottajan hallinnoiman verkkotunnuksen ja on muotoa:

```
http://{verkkotunnus}/doc/{aineistotunnus}/{paikallinen tunnus}
[//{versiotunnus}]
```

3.4 RDF ja RDF-tiedon sarjallistaminen

Kaikki linkitetty tieto webissä noudattaa RDF-tietomallia. RDF-tietomalli on suunnattu graafi, jossa jokainen solmun, linkin ja toisen solmun yhdistelmä muodostaa tietokolmikon eli tripletin (engl. triple). Yksi tripletti vastaa yhtä väittämää (engl. statement), jossa lähtösolmu edustaa subjektia, linkki predikaattia ja päätesolmu objektia. Subjektina toimiva solmupiste on joko resurssi, jolla on yksilöivä URI-tunnus, tai ns. tyhjä solmu (engl. blank node) eli anonyymi resurssi ilman URI-tunnusta. Objektina toimiva solmupiste voi olla resurssi tai pelkkä literaaliarvo. Solmupisteiden väliset linkit eli kolmikon predikaatit ovat aina resursseja, joilla on URI-tunnus. Linkit solmupisteiden välillä kuvaavat ominaisuuksia tai asioiden välisiä suhteita. Kuvassa 4 sinisellä on merkitty henkilöstä nimeltä Bob kertova epämuodollinen graafi, joka linkittyy muihin webin resursseihin. (W3C 2014a.)



Kuva 4: Epämuodollinen graafi, joka koostuu tietokolmikoista. Lähde: W3C 2014b.

RDF-tietomalli on suunniteltu mahdollistamaan eri rakenteisten ja useammasta eri lähteestä tulevien tietoaaineistojen yhdistämisen. Tietomallissa eri tietoaaineistoissa olevat yksittäiset tietoresurssit voidaan linkittää toisiinsa HTTP-muotoisten URI-tunnusten avulla. RDF-tietomalli ja URI-tunnusten käyttäminen resurssien identifiointiin mahdollistaa, että kuka tahansa voi viitata mihin tahansa maailmanlaajuisesti ja yksiselitteisesti. (Heath ym. 2011, s. 17–19.)

RDF-graafeja voidaan tallentaa joukkona tripletejä relaatiotietokantoihin tai erityisesti RDF-tiedolle suunniteltuihin RDF-tietokantoihin (engl. triplestore). RDF-tietokannoissa olennaista on, että tietoa voidaan hakea tietokantaa hyödyntävän SPARQL-palvelun (engl. SPARQL endpoint, kts. 3.5) avulla.

RDF-tietomallin mukainen tieto voidaan sarjallistaa useammassa eri sarjallistamismuodossa, jotka on suunniteltu erilaisiin tarpeisiin. Tämän työn yhteydessä luodussa prototyypipalvelussa RDF-tieto tarjotaan palvelun käyttäjille kolmessa eri sarjallistamismuodossa: RDF/XML, Turtle ja JSON-LD.

RDF/XML on W3C:n suositus (kts. W3C 2014c) alunperin vuodelta 1999 ja myös RDF:n alkuperäinen sarjallistamismuoto. RDF/XML on XML-muotoinen esitys RDF-tiedosta. XML-syntaksissa URI-tunnukset jaetaan kahteen osaan: nimiavaruusosaan (engl. namespace) ja paikallinen tunnus -osaan (engl. local name). Jako on siis hieman erilainen kuin URI-tunnuksen muodostamisprosessissa (kts. luku 3.2) käytetty. Resurssit voivat esiintyä joko XML-elementteinä tai XML-attribuutteina. RDF-tietomalli muodostuu RDF/XML-muodossa sisäkkäisistä elementeistä, jotka vastaavat aina vuorotellen solmua ja linkkiä. Tätä sarjallistamismuotoa pidetään ihmiselle haastavana tulkita ja muokata, joten muita sarjallistamismuotoja on kehitetty. (Heath ym. 2011, s. 20).

Turtle on kompakti tekstimuotoinen esitys RDF-tiedolle (Beckett ym. 2015). Se tukee nimiavaruuslyhenteiden (vastaavia kuin RDF/XML-tietoformaattissa) käyttöä ja on suunniteltu mahdollisimman ihmisluettavaksi (Heath ym. 2011, s. 21). Turtlessa RDF-kolmikot esitetään peräkkäisinä elementteinä niin, että yhdestä subjektista kertovat kolmikot voidaan ryhmittää peräkkäisiksi riveiksi. Kuvassa 5 on RDF-graafi esitetty Turtle-muodossa. RDF-graafi on saatu antamalla kuvan 4 graafin resursseille URI-tunnukset. Turtle-muotoisen esityksen alussa on määritelty esityksessä käytettävät nimiavaruuslyhenteet (engl. prefix), joihin viittamalla itse tiedon esityksestä saadaan huomattavasti kompaktimpi. Turtle-notaatiossa samaa subjektia tai predikaattia ei tarvitse toistaa joka rivillä, jolloin sen luettavuus pysyy hyvänä. Kuvassa 5 olevassa esimerkissä on esitetty neljä eri väittämää subjektista ”http://example.org/bob#me” (lyh. ex:bob#me), kaksi subjektista ”http://www.wikidata.org/entity/Q12418” (lyh. wd:Q12418) ja yksi subjektista ”http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619”. Turtle-muodon ja RDF/XML-muodon välillä voidaan toteuttaa automaattinen muunnos sarjallistamismuodosta toiseen.

Turtle-notaation nimiavaruuslyhennekäytäntöä käytetään myös tässä työssä esiteltävien sanastojen ja ontologioiden termien määrittelyjen ja esittelyjen yhteydessä. Li-

säksi kaikki työssä esitettävä RDF-tieto on Turtle-muodossa. Osasta Turtle-muotoisia RDF-kuvauksia on jätetty nimiavaruuslyhenteitä merkitsemättä. Kaikki käytetyt nimiavaruuslyhenteet löytyvät tämän työn Lyhenteet-osiosta (kts. sivu vi).

```
@prefix ex: <http://example.org/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix schema: <http://schema.org/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix wd: <http://www.wikidata.org/entity/> .

ex:bob#me
  a foaf:Person ;
  foaf:knows ex:alice ;
  schema:birthDate "1990-07-04"^^xsd:date ;
  foaf:topic_interest wd:Q12418 .

wd:Q12418
  dcterms:title "Mona Lisa" ;
  dcterms:creator
    <http://dbpedia.org/resource/Leonardo_da_Vinci> .

<http://data.europeana.eu/item/04802/
  243FA8618938F4117025F17A8B813C5F9AA4D619>
  dcterms:subject wd:Q12418 .
```

Kuva 5: Esimerkki Turtle-sarjallistamismuodosta. Lähde: W3C 2014b.

JSON-LD on RDF/XML- ja Turtle-notaatioihin verrattuna uusin sarjallistamismuoto. Se on kehitetty vastaamaan etenkin ohjelmistokehittäjien tarpeisiin luomalla siitä JSON-tietoformaatin (engl. JavaScript Object Notation) kanssa yhteensopiva linkitetyn tiedon formaatti (Sporny ym. 2014). Web-kehityksessä käytettävissä ohjelmointikielissä, kuten JavaScript-kielessä hyödynnettävälle JSON-syntaksin käsittelemiseksi on olemassa paljon valmiita työkaluja. Usein webistä saatava tieto on JSON-muodossa ja JSON-LD-sarjallistamismuodon yksi suunnitteluideoista on ollut, että JSON-muodossa oleva tieto olisi mahdollisimman yksinkertaisesti muunnettavissa linkitetyn tiedoksi JSON-LD-muotoon (Sporny ym. 2014).

JSON-tieto on joukko JSON-objekteja, jotka vastaavat JSON-LD-sarjallistamismuodossa linkitetyn tiedon resursseja. JSON-objekti koostuu aaltosulkeiden väliin määritellyistä avain-arvo-pareista. Avaimet ovat merkkijonoja ja JSON-LD-muodossa niiden täytyy olla yksikäsitteisiä resurssin sisällä. Jotta avain-arvo-pareihin saadaan liitettyä semantiikkaa, on JSON-LD-standardissa määritetty joukko avainsanoja, joilla on

tietty merkitys. Tärkeimpiä esimerkkejä niistä ovat ”@id”- ja ”@context” -avainsanat. ”@id”-avainsanalla annetaan resurssille URI-tunnus. ”@context”-avainsanalla voidaan määritellä termejä, joita käytetään resurssin kuvauksessa tai se voi olla URI-viittaus jossain muualla määritettyyn resurssin kuvauksessa käytettävään sanastoon (kts. kuva 6. (Sporny ym. 2014.)

Yksi JSON-LD-sarjallistamismuodon tärkeä ero verrattuna Turtle- ja RDF/XML-muotoihin, on mahdollisuus lisätä JSON-LD-muotoinen RDF-tieto suoraan HTML-dokumenttiin. Lisäys tehdään script-elementin sisälle, jolloin sovellukset voivat tulkita myös HTML-sivun RDF-tietona. (Sporny ym. 2014.)

3.5 SPARQL- ja GeoSPARQL -kyselykielet

RDF-tietoaaineisto voidaan jakaa resurssien URI-tunnusten mukaisten osoitteiden lisäksi myös SPARQL-palvelurajapinnan (engl. SPARQL endpoint) kautta. SPARQL-palvelusta haetaan tietoa SPARQL-kyselykielen mukaisilla kyselyillä. SPARQL-palvelu tulkitsee kyselyn ja palauttaa kyselyä vastaavat tiedot. SPARQL-kyselyillä voidaan monipuolisesti hakea kohdennettua tietoa RDF-tietograafista.

SPARQL-kyselykielellä on mahdollista tehdä kyselyitä, joissa RDF-graafista haetaan resursseja ja resursseihin liittyvää tietoa. SPARQL-kyselykieli mahdollistaa esimerkiksi loogiset unionit, erilaiset vastausten lajittelut, aggregaattitulokset ja sisäkkäiset kyselyt. SPARQL-kyselykieli tarjoaa kolme kyselytyyppiä: SELECT, ASK ja CONSTRUCT. SELECT-kyselyllä vastaus saadaan esimerkiksi taulukkomuodossa sisältäen haetut resurssien tiedot, literaaliarvot ja URI-tunnukset tai aggregaattitulokset. ASK-kysely palauttaa vastauksena totuusarvon, joka kertoo onko kyselylle olemassa ratkaisu. CONSTRUCT-kyselyn avulla voidaan kyselytulosten perusteella määritellä ja muodostaa uusi linkitetyn tiedon graafi. SPARQL-kyselyitä voidaan myös ketjuttaa (engl. Federated Query) käyttämään myös muita SPARQL-palveluita. Tämä mahdollistaa tietojen yhdistäminen eri tietoaaineistoista kyselyjen yhteydessä. (W3C 2013a.)

GeoSPARQL-standardi (kts. OGC 2012b) määrittelee SPARQL-kyselykielen laajennoksen, joka mahdollistaa spatiaalisiin operaatioihin perustuvat kyselyt linkitetyle tiedolle. Kuvassa 7 on esimerkki kyselystä, jossa haetaan kaikki kohteet, jotka sijaitsevat määritetyn polygonin sisällä. Standardissa määritellyt spatiaaliset operaatiot vastaavat OGC:n standardissa Simple Feature Access (kts. OGC 2005, tunnetaan myös nimellä ISO 19125) määriteltyjä spatiaalisia operaatioita *distance*, *buffer*, *convexHull*, *intersection*, *union*, *difference*, *symDifference*, *envelope* ja *boundary*. Spatiaalisten kyselyiden toteuttaminen edellyttää, että paikkatieto on mallinnettu GeoSPARQL-standardin määrittelemän ontologian mukaisesti. GeoSPARQL-ontologia esitellään luvussa 3.7.

```

{
  "@context": {
    "dbpedia": "http://dbpedia.org/resource/",
    "dcterms": "http://purl.org/dc/terms/",
    "ex": "http://example.org/",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "schema": "http://schema.org/",
    "wd": "http://www.wikidata.org/entity/",
    "xsd": "http://www.w3.org/2001/XMLSchema#"
  },
  "@graph": [{
    "@id": "ex:bob",
    "@type": "foaf:Person",
    "foaf:knows": {
      "@id": "ex:alice"
    },
    "foaf:topic_interest": {
      "@id": "wd:Q12418"
    },
    "schema:birthDate": {
      "@type": "xsd:date",
      "@value": "1990-07-04"
    }
  },
  {
    "@id": "http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619",
    "dcterms:subject": {
      "@id": "wd:Q12418"
    }
  },
  {
    "@id": "wd:Q12418",
    "dcterms:creator": {
      "@id": "dbpedia:Leonardo_da_Vinci"
    },
    "dcterms:title": "Mona Lisa"
  }
]}

```

Kuva 6: Esimerkki JSON-LD-sarjallistamismuodosta.

```

PREFIX my: <http://example.org/ApplicationSchema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

SELECT ?f
WHERE { ?f my:hasPointGeometry ?fGeom .
        ?fGeom geo:asWKT ?fWKT .
        FILTER (geof:sfWithin(?fWKT,
" <http://www.opengis.net/def/crs/OGC/1.3/CRS84>
        Polygon ((-83.4 34.0, -83.1 34.0,
                  -83.1 34.2, -83.4 34.2,
                  -83.4 34.0))"^^geo:wktLiteral))

```

Kuva 7: Esimerkki GeoSPARQL-kyselystä. Lähde: OGC 2012b.

3.6 RDF-skeema, RDF-sanastot ja OWL 2 -ontologiakieli

RDF-tietomalli tarjoaa tavan luoda väittämiä liittyen tietoresursseihin. Se ei kuitenkaan sisällä itsessään minkäälaisista tietoa siitä, mitä nämä resurssit ovat. Tarvitaan erillisiä sanastoja, joissa voidaan määritellä semantiikkaa käytettäville resursseille. Tällaisten sanastojen kehittämiseksi on luotu RDF-skeema (lyh. RDFS, engl. RDF Schema), joka tarjoaa perustan RDF-tiedon semanttiselle kuvaamiselle. RDF-skeema määrittelee luokka-käsitteen, jonka avulla resursseja voidaan kategorisoida. Resurssi voidaan määrittää kuuluvaksi, johonkin luokkaan ominaisuuskäsitteellä ”rdf:type” (lyhennetty Turtle-syntaksissa merkillä ”a” kuvassa 5). Luokka- ja ominaisuuskäsitteiden lisäksi RDF-skeemassa määritellään käsitteet mm. luokka- ja ominaisuushierarkian, ominaisuuksien arvo- (engl. range) ja sovellusalueiden (engl. domain) sekä luokkien instanssien kuvaamiseksi (kts. taulukot 1 ja 2). ”rdfs:Class”-, ”rdfs:Property”- ja ”rdfs:Datatype”-luokkia käytetään, kun halutaan määritellä uusia luokkia, ominaisuuksia tai tietotyyppisiä. (W3C 2014b.)

Taulukko 1: RDF-skeeman määrittelemiä luokkia

Luokka	kuvaus
rdfs:Resource	Luokka resursseille, kaikelle.
rdfs:literal	Literaaliarvojen luokka.
rdf:langString	Kielimääritteen sisältävien literaaliarvojen luokka.
rdfs:Class	Luokkien luokka.
rdf:Property	Ominaisuuksien luokka.
rdfs:Datatype	Tietotyyppien luokka.

RDF-sanastot ja ontologiat ovat osa RDF-tietomallia. Semanttisen webin viitekehysessä ne ovat toistensa synonyymejä, mutta usein vain laajemmista ja monimut-

Taulukko 2: RDF-skeeman määrittelemiä ominaisuuksia

Ominaisuus	kuvaus	domain	range
rdf:type	Subjekti on luokan instanssi.	rdfs:Resource	rdfs:Class
rdfs:subClassOf	Subjekti on luokan aliluokka.	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	Subjekti on ominaisuuden alio-minaisuus.	rdf:Property	rdf:Property
rdfs:domain	Ominaisuussubjektin sovellus-alue.	rdf:Property	rdfs:Class
rdfs:range	Ominaisuussubjektin arvo-alue.	rdf:Property	rdfs:Class
rdfs:label	Subjektiresurssin nimi.	rdfs:Resource	rdfs:Literal
rdfs:comment	Subjektiresurssin kuvaus.	rdfs:Resource	rdfs:Literal
rdfs:seeAlso	Lisätietoa subjektiresurssista.	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	Subjektiresurssin määritelmä.	rdfs:Resource	rdfs:Resource

kaisemmista sanastoista käytetään nimitystä ontologia. Sanastot eli tietosisältöä määrittelevät termit ovat RDF-tietomallissa samanlaisia resursseja kuin itse tietosisältöresurssit. Näin ollen metatiedon lisääminen tietoaineistoon onnistuu samoilla tekniikoilla kuin itse tiedon lisääminen. RDF-skeeman, kuten myös RDF-skeemaan perustuvan OWL-ontologiakielen avulla, voidaan tiedolle määritellä halutunlainen rakenne.

Yleisiä RDF-sanastoja ovat esimerkiksi Dublin Core, schema.org, VoID ja SKOS. Dublin Core (kts. DCMI Usage Board 2015) määrittelee metatiedon merkitsemiseksi käytettävän sanaston DCMI Metadata Terms, joka sisältää mm. tietoaineistoa kuvaavat ominaisuudet: tekijä, julkaisija ja otsikko. Schema.org (kts. schema.org 2015) -sanasto on suurten hakukoneyhtiöiden, kuten Googlen, Yahoön ja Yandexin hyödyntämä yhteisöllisesti kehitettävä sanasto. Web-kehittäjät voivat merkitä webissä julkaistavaa tietoa sanaston avulla, jolloin hakukoneet osaavat tulkita tietoa ja tarjota sitä asiakkailleen kohdennetummin. Simple Knowledge Organization System (lyh. SKOS, kts. W3C 2009) on erilaisten asiasanastojen julkaisemiseksi määritelty sanasto. Se on W3C:n suosittelu tapa julkaista olemassa olevaa rakenteista tietoa linkitettyinä tietona. SKOS-ontologia voidaan katsoa olevan myös oma ontologiakielensä, koska se määrittelee tavan ilmaista asioiden suhteita ja sen avulla voidaan rakentaa toisia rakenteellisia sanastoja. (W3C 2014b.)

Yksi tässä työssä hyödynnettävistä RDF-sanastoista on Vocabulary of Interlinked Datasets (lyh. VoID). VoID on sanasto RDF-muotoisten tietoaineistojen kuvaamiseksi. VoID määrittelee käsitteen tietoaineisto, jolla on käyttötarkoitus, aihe ja ylläpitäjä. Tietoaineistoa kuvaava resurssi on "void:Dataset". Sanastolla voidaan ilmaista tietoaineiston olevan tarjolla määritetyn URIn kautta linkkien välityksellä, määritetyn SPARQL-palvelurajapinnan kautta tai yhtenä tiedostona määritetysä URI-osoitteessa. Jos aineistossa on jokin kohde, josta linkkien avulla päästään

käsiksi kaikkiin muihin kohteisiin, niin tämä suhde voidaan ilmaista predikaatilla ”void:rootResource”. Tietoaineistoon liittyvän metatiedon ilmaisemiseksi hyödynetään Dublin Core -sanastoa. VoID-sanaston avulla voidaan tietoaineistolle yksinkertaisesti määrittää myös käyttölisenssi, kertoa aineistontarjoajasta ja kuvailla tietoaineistoa. VoID-sanaston avulla voidaan ilmaista, missä sarjallistamismuodoissa aineisto tarjotaan.

W3C:n OWL 2 -ontologiakieli on kehitetty monimutkaisen ja rikkaan tiedon jäsentämiseksi. OWL 2 on huomattavasti ilmaisuvoimaisempi kuin RDF-skeema ja SKOS. OWL 2 -ontologiakielellä voidaan esittää tietoa asioista, niiden ryhmistä ja suhteista toisiin asioihin. OWL 2 -ontologiakieli tarjoaa työkalun ontologioiden formaaliin määrittelyyn. OWL 2 on deklarativinen kieli, joka kuvailee asioiden suhteet loogisesti. Tämä mahdollistaa, että myös koneet voivat ymmärtää tietoa ja tehdä loogisia päätelmiä uuden tiedon saamiseksi olemassaolevaan tietoon perustuen. OWL 2 on aikaisemman OWL-ontologiakielen (myöh. OWL 1 -ontologiakieli) laajennos. Kaikki OWL 1 -ontologiat ovat myös valideja OWL 2 ontologioita. (W3C 2012b.)

OWL 2 -ontologiakieli hyödyntää avoimen maailman oletusta (engl. Open World Assumption). Toisin kuin mahdollisesti suljettujen tietokantojen tapauksissa, OWL 2 -ontologiakielen mukaisessa tiedossa puuttuvasta tiedosta ei voida päätellä, että jokin asia on epätosi. OWL 2 ei tarjoa mahdollisuutta ilmaista, mitä tietoja jostain resursseista tulisi tarjota. Tämä on tärkeä tiedostaa muunnettaessa tietomallia ontologiakielelle. (W3C 2012b.)

OWL 2 -ontologiakielelle on luotu kaksi erilaista formaalia semantiikka, jotka määrittelevät päättelysäännöt lisätiedon johtamiseksi OWL-ontologioista. OWL 2 ei kuitenkaan määrittele, kuinka mahdollinen päättely tulee toteuttaa käytännössä. Ensimmäinen formaali semantiikka on Direct Semantics (kts. W3C 2012a), jota voidaan hyödyntää OWL 2 -ontologiakielen rajatun osakielen OWL 2 DL (Description Logics) mukaisille ontologioille. OWL 2 DL on kehitetty syntaksiltaan tarkemmaksi, mikä mahdollistaa täydellisen päättelyn sellaisille ontologioille, jotka toteuttavat OWL 2 DL -osakielen vaatimukset. Toinen formaali semantiikka on RDF-Based Semantics (kts. W3C 2012c), jota voidaan hyödyntää kaikille OWL-ontologioille. RDF-Based Semantics on RDF-tietomallille ja RDF-skeemalle määritetyn formaalin semantiikan laajennos, jolloin OWL 2 -ontologiat voidaan ajatella lähtökohtaisesti RDF-graafina. (W3C 2012b.)

Keskeisiä käsitteitä tiedon mallintamisessa OWL 2 -ontologiakielellä ovat aksioomat (engl. axioms), entiteetit (engl. Entities) ja ilmaisut (engl. Expressions). OWL-ontologiassa esitetyt väittämät ovat aksioomia, jotka ovat ontologian mukaan tosia. Entiteetit ovat joko luokkia, luokkien yksilöitä (engl. individuals) tai ominaisuuksia. Ominaisuudet on jaettu kohdeominaisuuksiin (engl. object property), joissa yksilöllä on jokin suhde toiseen yksilöön ja tietotyyppiominaisuuksiin (engl. datatype property), joissa yksilöllä on jokin suhde jonkin tietotyyppin mukaiseen arvoon. Kolmas ominaisuustyyppi seliteominaisuus (engl. annotation property) mahdollistaa ontologian tai sen osien kuvailun. Yksi keskeisimmistä OWL 2 -ontologiakielen

ominaisuuksista on sen ilmaisut, jotka mahdollistavat uusien entiteettien määrittämisen olemassa olevia entiteettejä yhdistämällä. (W3C 2012b.) Jatkossa OWL 2 -ontologiakielestä käytetään tässä työssä yleisesti nimitystä OWL-ontologiakieli.

3.7 GeoSPARQL-ontologia

Paikkatiedon kuvaamisen linkitettyinä tietona mahdollistavan ontologian luominen on paikkatiedon erityisyyden takia monimutkainen tehtävä. Paikkatieto on aina moniulotteista tietoa, jota voidaan mallintaa hyvin monilla eri tavoin. Paikkatiedon kuvaamista mahdollistavia sanastoja onkin kehitetty useampia 2000-luvulla (Battle ym. 2012, s. 358). W3C:n Semantic Web Interest Group -työryhmä loi vuonna 2003 paikkatietoa kuvaavan Basic Geo (WGS84 lat/long) Vocabulary - RDF-sanaston (kts. W3C 2006b). Basic Geo -sanasto mahdollistaa pistegeometrian WGS84 - maantieteellisten koordinaattien sekä -korkeuskoordinaatin lisäämisen RDF-tietoon. Vuonna 2007 W3C:n Geospatial Incubator Group -työryhmä sai valmiiksi Geo OWL -ontologian (kts. W3C 2007), joka mahdollistaa GML-merkintäkielen mukaisten piste-, viiva- ja aluegeometrioiden ilmaisemisen RDF-tietona. Geo OWL -ontologia ei myöskään mahdollista muita koordinaatistoja kuin WGS84-koordinaatit. (Battle ym. 2012, s. 358).

NeoGeo-sanasto (kts. Norton ym. 2012) perustuu GML Simple Features -profiiliin ja tarjoaa myös Region Connection Calculus (lyh. RCC) -logiikkaan perustuvat RCC8-topologiasuhteet. NeoGeo-sanastossa on tehty valinta, että yksittäiset koordinaattipisteet esitetään resurssina, eikä literaaliarvona. Tästä on se etu, että saman URIn perusteella voidaan tarjota koordinaattipiste halutussa muodossa Content Negotiation -tekniikan avulla. Koordinaattipisteitä kuvaavat resurssit tekevät mallista kuitenkin hankalasti hyödynnettävän. Esimerkiksi pistegeometrioista koostuvaa geometriaa esittävät RDF-kokoelmat ovat haastavia SPARQL-kyselyissä. Lisäksi yksittäiselle koordinaattipisteelle annettu URI-tunnus ei tarjoa juuri muuta lisäarvoa kuin sen eri esitysmuodot. (Battle ym. 2012, s. 360.)

GeoSPARQL-standardi (OGC 2012b) on OGC:n yritys tarjota yhteinen tapa julkaista ja kysellä paikkatietoa RDF-tietona. GeoSPARQL mahdollistaa monipuoliset spatiaaliset kyselyt ja eri koordinaatistojen sekä topologiasuhteiden ilmaisemisen. Standardi koostuu kuudesta komponentista:

- 1 Ydinkomponentti, joka määrittelee ylätasen RDFS/OWL -luokat paikkatietokohteille.
- 2 Topologiasanasto, joka määrittelee RDF-ominaisuudet paikkatietokohteiden topologisten suhteiden määrittämiseksi ja kyselemiseksi.
- 3 Geometriakomponentti, joka määrittelee RDFS-tietotyyppit geometrian sarjallistamiseksi, geometrioihin liittyvät RDF-ominaisuudet ja geometriaan perustuvat spatiaaliset funktiot.

- 4 Topologiakomponentti, joka määrittelee topologiaan perustuvat spatiaaliset funktiot.
- 5 RDFS Entailment Extension -komponentti, joka määrittelee GeoSPARQL-toteutuksessa tuetut topologiamallit ja geometrioiden sarjallistamismuodot.
- 6 Query Rewrite Extension -komponentti, joka määrittelee GeoSPARQL-toteutuksessa käytetyt säännöt kahden kohteen välisen topologiakyselyn muuntamiseksi koskemaan kohteiden geometriaa.

Ydinkomponentissa määritellään kaksi luokkaa: "geo:SpatialObject" (kts. Kuva 8) ja "geo:Feature" (kts. kuva 9). Geometriakomponentti määrittelee "geo:Geometry"-luokan (kts. kuva 10). Luokat "geo:Geometry" ja "geo:Feature" ovat luokan "geo:SpatialObject" aliluokkia. Paikkatietokohde ("geo:Feature") ja geometria ("geo:Geometry") ovat erotettu toisistaan käsitteellä "owl:disjointWith". Luokka "geo:Geometry" vastaa määritelmän mukaan ISO 19107 -standardin GM_Object UML-mallia. Luokka "geo:Feature" vastaa määritelmän mukaan ISO 19109 -standardin määrittelemän abstraktin General Feature Model -kohdemallin (lyh. GFM) GF_Feature -kohteen instanssijoukkoa GFI_Feature. GFI_Feature on määritelty ISO 19156 -standardissa. GFM-kohdemallissa geometria on yksi kohteen ominaisuuksista ja GeoSPARQL-ontologiassa tämä suhde ilmaistaan "geo:hasGeometry"-predikaatilla (kts. kuva 11).

```
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

geo:SpatialObject a rdfs:Class,
                  owl:Class;
  rdfs:isDefinedBy <http://www.opengis.net/spec/geosparql/1.0>;
  rdfs:label       "Spatial Object"@en;
  rdfs:comment     "The class Spatial Object represents everything
                    that can have a spatial representation. It is
                    superclass of feature and geometry"@en .
```

Kuva 8: Luokan "geo:SpatialObject" määritelmä. Lähde: OGC 2012b.

Topologiasanasto määrittelee kolme erilaista topologiamallia (engl. relation family): OGC:n Simple Features, RCC8:n ja Egenhofer:n mukaiset mallit. GeoSPARQL-standardin OGC:n Simple Features -mallin mukaiset topologiset suhteet ja niiden URI-tunnukset on esitetty taulukossa 3. Näille suhteille löytyy vastaavuudet myös RCC8- ja Egenhofer-topologiamalleista. Topologiasuhteet voidaan standardin mukaan yhdistää sekä kohteisiin ("geo:Feature") että geometrioihin ("geo:Geometry").

GeoSPARQL-standardin geometriakomponentissa määritellään erilaisia ominaisuuksia, joilla geometrioita voidaan kuvata. Geometrian ulottuvuuksia voidaan kuvata

```

geo:Feature a rdfs:Class,
    owl:Class;
rdfs:isDefinedBy <http://www.opengis.net/spec/geosparql/1.0>;
rdfs:label      "Feature"@en;
rdfs:subClassOf geo:SpatialObject;
owl:disjointWith geo:Geometry;
rdfs:comment    "This class represents the top-level feature type.
                  This class is equivalent to GFI_Feature defined
                  in ISO 19156, and it is superclass of all feature
                  types."@en .

```

Kuva 9: Luokan "geo:Feature" määritelmä. Lähde: OGC 2012b.

```

geo:Geometry a rdfs:Class,
    owl:Class;
rdfs:isDefinedBy <http://www.opengis.net/spec/geosparql/1.0>;
rdfs:label      "Geometry"@en;
rdfs:subClassOf geo:SpatialObject;
owl:disjointWith geo:Feature;
rdfs:comment    "The class represents the top-level geometry
                  type. This class is equivalent to the UML
                  class GM_Object defined in ISO 19107, and it
                  is superclass of all geometry types."@en .

```

Kuva 10: Luokan "geo:Geometry" määritelmä. Lähde: OGC 2012b.

tietotyyppiominaisuuksilla "geo:dimension", "geo:coordinateDimension" ja "geo:spatialDimension". Itse geometria esitetään OGC:n Well-known Text (lyh. WKT, kts. OGC 2005) -muodossa tai GML-muodossa. Niille määritellyt RDFS-tietotyypit ovat "geo:wktLiteral" ja "geo:gmlLiteral". Geometrian sarjallistuksiin viitataan predikaatilla "geo:asWKT" tai "geo:asGML", jotka ovat tietotyyppiominaisuuden "geo:hasSerial-

```

geo:hasGeometry a rdf:Property,
    owl:ObjectProperty;
rdfs:isDefinedBy <http://www.opengis.net/spec/geosparql/1.0>;
rdfs:label      "has Geometry"@en;
rdfs:comment    "A spatial representation for a given feature."@en;
rdfs:domain     geo:Feature;
rdfs:range      geo:Geometry .

```

Kuva 11: Predikaatin "geo:hasGeometry" määritelmä. Lähde: OGC 2012b.

Taulukko 3: Simple Features -mallin topologiset suhteet. Lähde: OGC 2012b.

Suhteen nimi	Suhteen URI
equals	geo:sfEquals
disjoint	geo:sfDisjoint
intersects	geo:sfIntersects
touches	geo:sfTouches
within	geo:sfWithin
contains	geo:sfContains
overlaps	geo:sfOverlaps
crosses	geo:sfCrosses

lization aliominaisuuksia”. WKT-muoto mahdollistaa ainoastaan OGC:n Simple Features -mallin mukaisten geometrioiden ilmaisemisen, GML-muoto mahdollistaa myös monimutkaisemmat geometriat. Sekä GML-geometrioille että Simple Features -geometrioille löytyy omat RDF-sanastot, joita käytetään GeoSPARQL-ontologiassa.

4 Aiemmat tutkimukset

Schade ja Cox (2010) tuovat tutkimuksessaan esiin tarpeen paikkatietoaineistojen laajemmalle jakamiselle ja sen yhdistettävyydelle toisiin aineistoihin. Paikkatietoinfrastruktuureja ohjaa monet standardit, joita tuottavat mm. OGC, ISO/TC 211 Geographic information/Geomatics ja INSPIRE. Niitä käytetään käytännössä vaihtelevasti ja soveltaen. Näin ollen aineistojen yhteensovittaminen on haastavaa. Aineistojen saattamiseksi yhteentoimivaksi on ehdotettu mm. linkitetyn tiedon tekniikoita. GML-tietomalli on luonteeltaan yhteneväinen RDF-tietomallin kanssa. GML-tietomalli mahdollistaa kohteeseen liittyvien resurssien merkitsemisen sisäkkäisesti tai linkittämällä käyttäen W3C:n XLink-mekanismia. GML-muotoisen tiedon kohde-ominaisuus-rakenne ja XLink-linkitys ovat suoraan muunnettavissa esim. RDF/XML-muotoon. Edellytyksenä on, että GML-muotoisessa tiedossa käytetään URI-muotoisia tunnuksia. Tutkimuksessa ehdotetaan, että lisäämällä GML-muotoista dataa tarjoavaan palveluun Content Negotiation -ominaisuus, ovat nykyiset paikkatietoinfrastuktuurit yksinkertaisen muunnoksen avulla tarjottavissa osana semanttista webiä. (Schade ja Cox 2010.)

Schaden ja Coxin (2010) ehdottama menetelmä tarjoaa yleisen ratkaisun paikkatiedon tarjoamiseksi linkitettynä tietona. Tässä työssä GML-RDF/XML -muunnoksen lisäksi muokataan paikkatietoaineiston tietomallia, luodaan PNR-ontologia hyödyntäen GeoSPARQL-ontologiaa ja asetetaan paikkatietokohteet saataville URI-tunnusten mukaisista osoitteista.

Tschirner ym. (2011) esittelevät SPARQL-palvelun, joka mahdollistaa INSPIRE-yhteensopivan WFS-kohdepalvelun hyödyntämisen kohteiden tarjoamiseksi semanttisessa muodossa. Ideana on luoda INSPIRE-teeman GML-mallia vastaava OWL-ontologia, jota hyödynnetään palveluun syötettävässä SPARQL-kyselyssä. Palvelu muuntaa SPARQL-kyselyn OGC:n Filter Encoding -standardin mukaiset rajausehdot sisältäväksi WFS-kyselyksi ja hakee sillä kohteita WFS-kohdepalvelusta. Kohteet muunnetaan luodun OWL-ontologian mukaiseksi RDF-graafiksi, josta alkuperäisellä SPARQL-kyselyllä saadaan tavoiteltu vastaus. Esitetyn palveluarkkitehtuurin etuna on datan ylläpitäminen ainoastaan yhdessä paikassa, jolloin data on aina mahdollisimman ajantasaista. Myöskään INSPIRE-datan tarjoamisesta vastaavien viranomaisten ei tarvitse tarjota INSPIRE-yhteensopivan WFS-kohdepalvelun lisäksi mitään ylimääräistä, jotta kuvatus kaltainen SPARQL-palvelu voidaan toteuttaa. (Tschirner ym. 2011.)

Tschirnerin ym:n (2011, s. 79) tutkimuksessa on kehitetty säännöt UML-luokkadiagrammin muuntamiseksi OWL-ontologiaksi. Perussäännöt ovat:

1. UML-luokista tehdään OWL-luokkia.
2. Kaikista koodilistoista (engl. codelist) ja arvojoukoista (engl. enumeration) tehdään OWL-luokkia ja niiden arvoista OWL-yksilöitä eli koodilistaa tai arvojoukkoa vastaavan OWL-luokan jäseniä.

3. UML-attribuuteista tehdään OWL-kohdeominaisuuksia tai OWL-tietotyyppiominaisuuksia.
4. UML-assosiaatioista tehdään OWL-kohdeominaisuuksia.

Tässä tutkimuksessa hyödynnetään kyseistä säännöstöä Paikka-ontologian luomisessa luvussa 5.2.

Jones ym. (2014) esittelevät tutkimuksessaan LOD2WFS-sovelluksen, joka toimii perusajatukseltaan päinvastoin kuin Tschirnerin ym:n (2011) esittelemä SPARQL-palvelu. Siinä linkitetty avoin tieto (engl. Linked Open Data, lyh. LOD) integroidaan osaksi paikkatietoinfrastruktuuria tarjoamalla tieto WFS-kohdepalvelun kautta. Palvelu kuuntelee asiakasohjelman lähettämiä WFS-kyselyitä ja muuntaa ne SPARQL-kyselyiksi, joka välitetään SPARQL-palvelulle. SPARQL-palvelun palauttama RDF-tulosjoukko muunnetaan WFS/XML-dokumentiksi, joka palautetaan asiakasohjelmalle. Tämä mahdollistaa linkitetyn tiedon käsittelyn paikkatiedon tehokkaaseen analysoimiseen tarkoitetuilla GIS-ohjelmistoilla, jotka osaavat hyödyntää WFS-rajapintoja. (Jones ym. 2014.)

Norjan kansallisen karttalaitoksen (Kartverket) pilotissa on testattu paikkatietoaineiston tarjoamista linkitettyä tietona. Pilotissa luotiin ontologia INSPIRE-direktiivin mukaisille teemoille Hallinnolliset yksiköt (AU) ja Paikannimet (GN). Data tarjottiin RDF-tietona hyödyntäen Parliament TripleStore/SPARQL -palvelua. Pilotissa tehtiin muutamia huomioita. Aineiston tarjoamisessa linkitettyä tietona on paljon potentiaalia. Lisäksi tietomallin yksinkertaistaminen voi olla hyödyllistä ontologisoinnin yhteydessä. GeoSPARQL-ontologia on suositeltava ontologia paikkatiedon esittämiseksi linkitettyä tietona, mutta spatiaaliset kyselyt ovat toistaiseksi valitettavan hitaita Parliamentilla sekä monilla muilla RDF-tiedolle luoduilla ratkaisulla. (Guldborg 2015.)

Palveluista, jotka tarjoavat paikannimiä linkitettyä tietona, kattavin on todennäköisesti GeoNames.org (kts. GeoNames 2015). GeoNames.org toimii avoimella lisenssillä ja tarjoaa yli 10 miljoonaa paikannimeä maailmanlaajuisesti. GeoNames.org hyödyntää mm. olemassa olevia avoimia aineistoja ja joukkoistamista. GeoNames.org-palvelussa oleva aineisto on saatavilla URI-tunnusten perusteella RDF-muodossa.

Maanmittauslaitoksen paikannimirekisteri on saatavilla linkitettyä tietona ainakin Linked Data Finland (lyh. LDF) -tutkimusprojektin yhtenä tuotoksena (Linked Data Finland 2015). LDF on luonut aineistoa varten oman ontologian, joka hyödyntää W3C:n Basic Geo -sanastoa paikkojen koordinaattitiedon tarjoamisessa. Paikkojen sijoittuminen kuntiin, maakuntiin tai lääneihin ilmaistaan kulttuuriperinnön tallennukseen ja luokitteluun tarkoitettua CIDOC CRM -sanastoa (CIDOC 2015) hyödyntäen. Aineisto on tarjolla LDF:n tarjoaman oman aineistonselauksovelluksen kautta sekä SPARQL-palveluna.

5 Paikannimet linkitettynä tietona -palvelu

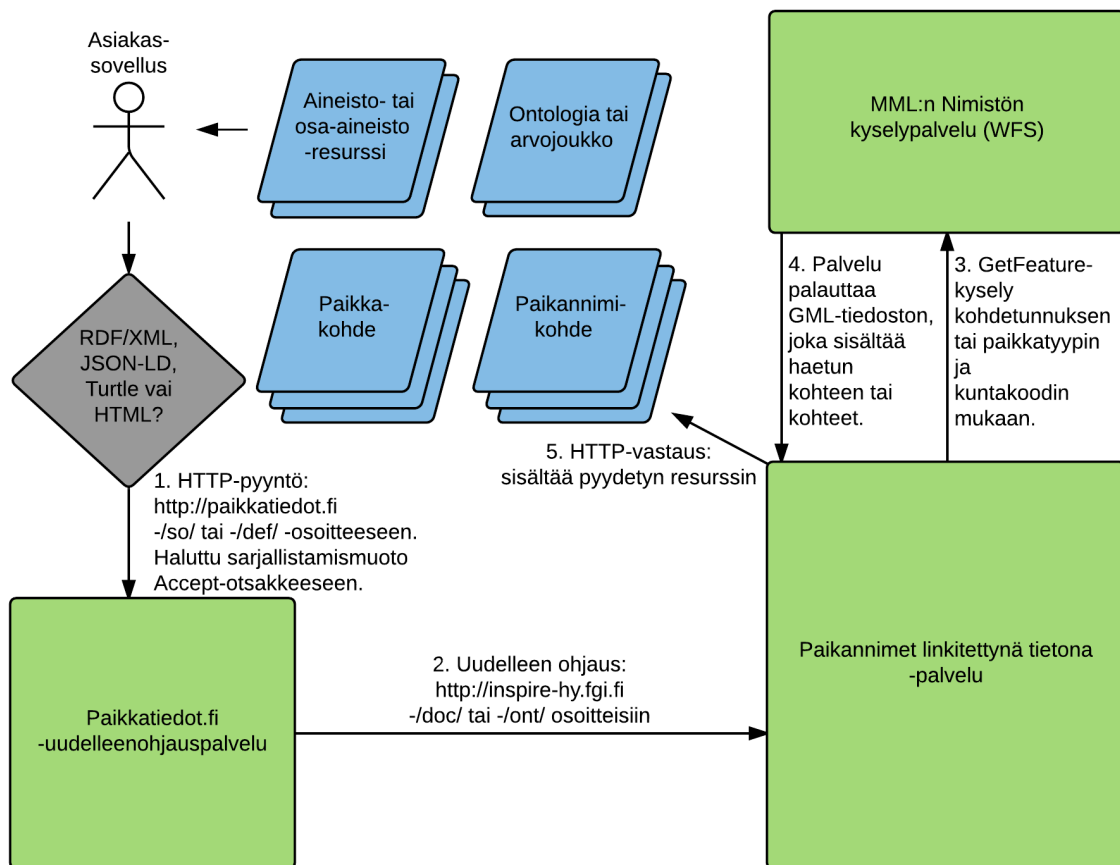
Osana tätä diplomityötä on luotu verkkopalvelu nimeltä Paikannimet linkitettynä tietona (kts. Hietanen 2015). Palvelu tarjoaa Maanmittauslaitoksen paikannimirekisterin linkitettynä tietona kohteiden URI-tunnuksia hyödyntämällä. Palvelu on prototyyppi, jonka avulla voidaan selvittää yleisesti tällaisen palvelun hyötyjä ja käyttötapauksia. Luotua toteutusta voidaan myös hyödyntää vastaavan tuotantopalvelun suunnittelussa. Prototyypin kehittämisen yhteydessä PNR:n paikkatietokohteille on annettu yksilöivät URI-tunnukset JHS 193 ”Paikkatiedon yksilöivät tunnukset”-suosituksen (JUHTA 2015) mukaisesti. Tietoaineiston tarjoamiseksi linkitettynä tietona PNR-tietomallille on luotu oma ontologia hyödyntäen olemassa olevia ontologioita, kuten GeoSPARQL-ontologiaa. Tietoaineiston kuvaamiseksi, aineiston jakamiseksi osa-aineistoihin ja aineiston kohteiden linkittämiseksi osa-aineistojen kautta toisiinsa on käytetty VoID-sanastoa.

Palvelun yleinen toimintaperiaate on esitetty kuvassa 12. Idea on, että ensimmäisessä vaiheessa asiakasohjelma tekee HTTP-pyynnön PNR:n aineistoresurssin, Paikka- tai Paikannimi-kohteen, ontologian tai arvojoukon URI-tunnuksen mukaiseen osoitteeseen. HTTP-pyynnön Accept-otsakkeella ilmoitetaan palvelulle haluttu esitysmuoto. Pyynnöt ohjautuvat Paikkatiedot.fi -uudelleenohjauspalvelusta Paikannimet linkitettynä tietona -palveluun (2. vaihe). Paikannimet linkitettynä tietona -palvelu muodostaa URI-tunnisteen perusteella tarvittavan WFS-kyselyn MML:n Nimistön kyselypalveluun (3. vaihe). Nimistön kyselypalvelu palauttaa haetun paikkatietokohteen tai -kohteet GML-muodossa (4. vaihe). Palvelu muokkaa tiedosta RDF-tietomallin ja PNR-ontologian mukaisen esityksen. Viimeisessä (5.) vaiheessa tieto palautetaan HTTP-vastauksen sisältönä asiakasohjelman pyytämässä sarjallistamis- tai esitysmuodossa.

Tässä luvussa esitellään palvelun lähtöaineisto (kts. luku 5.1), ontologian luominen aineistolle (kts. luku 5.2) sekä itse verkkopalvelun toteutus (kts. luku 5.3). Ontologian luomisen ja palvelun toteutuksen kuvailun yhteydessä pohditaan tehtyjen ratkaisujen perusteita. Näin pyritään vastaamaan tutkimuskysymyksiin siitä, miten paikkatietoaineiston tietomalli voidaan toteuttaa ontologiana ja kuinka aineisto voidaan tarjota linkitettynä tietona. Niiden jälkeen luvussa 5.4 pohditaan palvelun hyödyllisyyttä erilaisten mahdollisten käyttötapauksien avulla. Lopuksi luvussa 5.5 arvioidaan käytettyjen ratkaisujen toimivuutta ja esitetään palveluun liittyviä jatkokehitysajatuksia.

5.1 Paikannimirekisteri

Maanmittauslaitoksen nimistörekisteri koostuu paikannimirekisteristä (lyh. PNR) ja karttanimirekisteristä. PNR sisältää yli 800000 paikkaa, joihin liittyy yksi tai useampi paikannimi. Suomessa on viisi virallista kieltä. Suomen- ja ruotsinkieli ovat virallisia valtion kieliä. Lisäksi Suomen neljässä pohjoisimmassa kunnassa kolmella eri



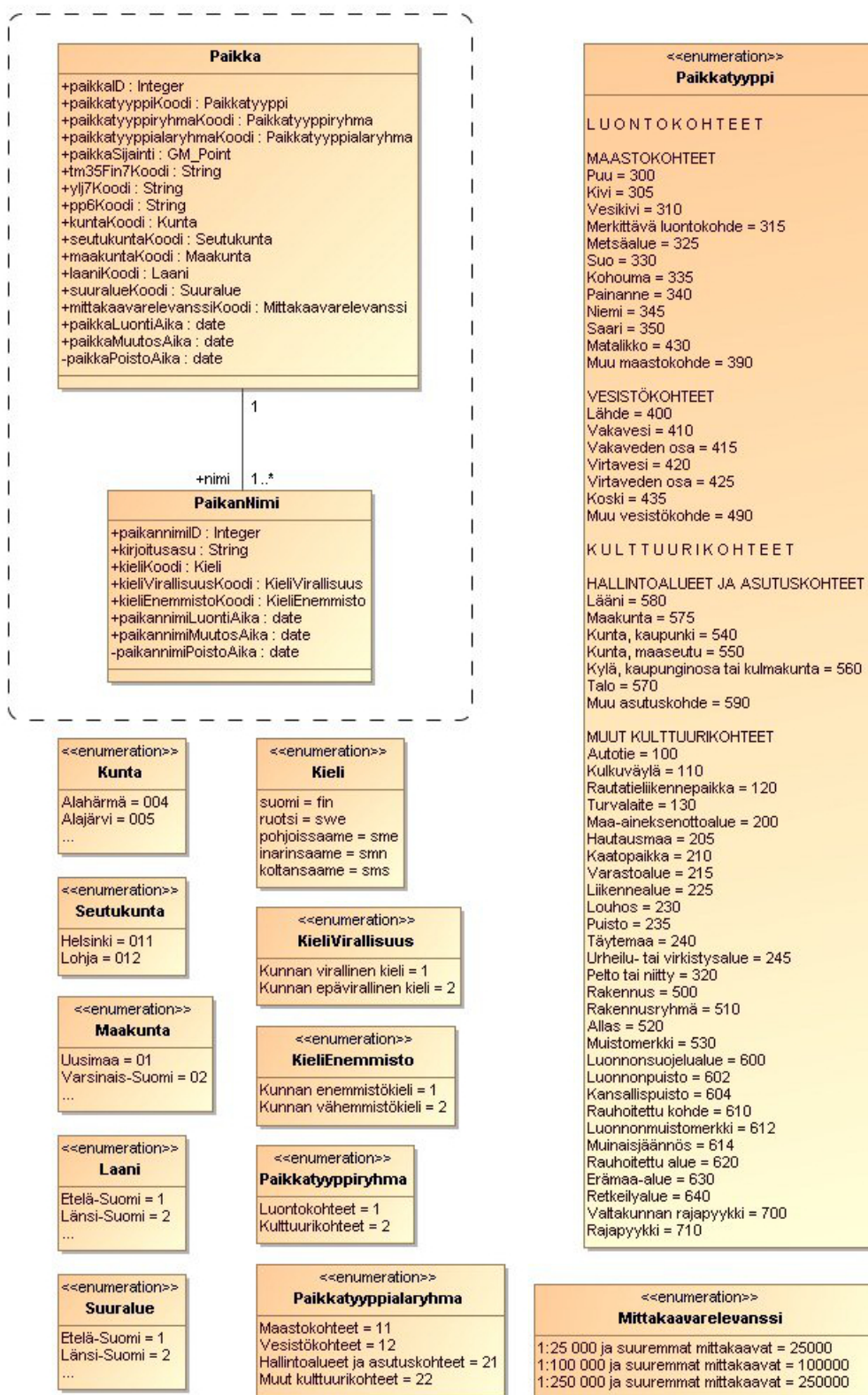
Kuva 12: Paikannimet linkitettyinä tietona -palvelun yleinen toimintaperiaate.

saamenkielellä: inarinsaamalla, koltansaamalla ja pohjoissaamalla, on virallinen asema. Näin yhdellä paikalla voi olla yhdestä viiteen erikielistä nimeä. On myös paikkoja, joilla on useampi kuin yksi samankielinen nimi. Nimistörekisteri on Suomen virallinen paikannimirekisteri. Sitä hyödynnetään Maanmittauslaitoksen karttatuotteissa sekä monenlaisessa selkeää ja tarkkaa paikannimitietoa vaativassa viestinnässä. (Leskinen 2015.)

Maanmittauslaitoksen Nimistön kyselypalvelu (WFS) sisältää paikannimirekisterikyselyjä varten kaksi rajapintatuotetta, Paikat ja Paikannimet. Tuotteiden tietosisältö tarjotaan GML-muodossa ja vaikka molemmilla tuotteiden tietosisältö on sama, niiden GML-sovellusskeemat ovat erilaisia. Sovellusskeemassa paikalle määritellään paikkaan liittyvät ominaisuustiedot, kuten sijainti, kuntakoodi ja paikkatyyppi. Lisäksi sille määritellään yksi tai useampi Paikannimi, jolle on määritelty omat ominaisuustiedot, kuten kirjoitusasu, kielikoodi ja lähde. Paikannimeä kuvaava GML-rakenne koostuu itse paikannimen ominaisuustiedoista, siihen liittyvän paikan ominaisuustiedoista sekä paikkaan liittyvien rinnakkaisnimien ominaisuustiedoista. Paikkoihin ja paikannimiin liittyvät tiedot löytyvät Paikat-tuotteen GML-esityksen rakennetta kuvaavasta UML-luokkadiagrammista (kts. kuva 13). (MML 2015.) UML-luokkadiagrammissa Paikan ja Paikannimen välillä on assosiaatiosuhde, jolloin GML-muodossa Paikkojen

ja Paikannimien suhde voitaisiin toteuttaa myös GML-merkintäkielen hyödyntämällä XLink-suhteella (OGC 2012a). MML:ssa on kuitenkin päädytty ratkaisuun tarjota kaikki kohteeseen (Paikka tai Paikannimi) liittyvät tiedot kohteeseen sisällytettyinä elementteinä. Tällöin käyttäjän ei tarvitse hakea palvelusta erikseen paikan ja paikkaan liittyvän nimen tietoja.

Paikat-tuotteen UML-luokkadiagrammissa (kts. kuva 13) on määritelty kaksi luokkaa Paikka ja Paikannimi. Paikka-kohteen sijainnin arvoalueeksi on määritelty GML-skeeman mukainen pistegeometria. Pistegeometrian oletuskoordinaatisto on ETRS-TM35FIN ja WFS-kohdepalvelun kautta kohteet voidaan hakea halutussa koordinaatistossa. Muiden Paikka- ja Paikannimi-luokan ominaisuuksien arvolueena on joko XML-skeemakielen mukainen alkeistyyppi, kuten "xsd:dateTime" tai "xsd:string" (kts. W3C 2012d), tai UML-kaaviossa määriteltyt arvojoukot (engl. enumeration lists), kuten Seutukunta tai Paikkatyyppi.



Kuva 13: Paikkanimirekisterin Paikat -tuotteen UML-luokkadiagrammi. Lähde: MML 2015.

5.2 PNR-ontologian luominen

Ontologisoinnissa on tärkeä erotella milloin lähtöaineiston tietomalli on kanonisoitu ja harmonisoitu tietomalli ja milloin taas ilmiötä vapaammin kuvaileva malli. Ensimmäisessä tapauksessa ontologisointi täytyy tehdä tarkasti noudattaen lähtöaineiston UML-tietomallia. Jälkimmäisessä tapauksessa voidaan mallia muokata linkitetyn datan periaatteiden avulla ja noudattaa ”Open World” -oletusta. (Cox 2013.) PNR-aineiston tietomalli on luotu ainoastaan kyseiselle aineistolle, joten ei ole tarkoituksenmukaista yrittää tehdä mallin täydellistä kopiota OWL-ontologiakielellä. Tämän työn yhtenä tarkoituksena on luoda linkitetyn tiedon edellyttämä ontologia, jolla PNR-aineiston tietomallin voi toteuttaa mahdollisimman selkeästi ja kattavasti. Esimerkiksi aineiston jakaminen hierarkkisesti osa-aineistoiksi voi helpottaa kokonaisuuden jäsentämistä, joten on kannattavaa hyödyntää OWL-ontologiakielen monipuolisia ominaisuuksia hierarkkisten suhteiden määrittämiseksi.

Paikat-tietotuotteen GML-rakennetta (kuvan 13 UML-kaavio ja tietotuotteen XML-skeema -tiedostot (MML 2015)) tutkimalla voidaan löytää ainakin kaksi erilaista tapaa jaotella aineistoa hierarkkisesti. Ensimmäinen tapa on paikkatyyppeihin perustuva jaottelu. Paikkatyyppeihin perustuvassa jaottelussa PNR:n paikat on jaettu tyypiltään kahteen paikkatyyppiryhmään: Luontokohteet ja Kulttuurikohteet. Paikkatyyppiryhmät on jaettu seuraaviin paikkatyypialaryhmiin:

- Maastokohteet,
- Vesistökohteet,
- Hallintoalueet ja asutuskohteet sekä
- Muut kulttuurikohteet.

Paikkatyypialaryhmistä kaksi ensimmäistä kuuluvat Luontokohteisiin ja kaksi jälkimmäistä Kulttuurikohteisiin. Erilaisia paikkatyyppejä, jotka on jaoteltu kuuluvaksi tasan yhteen paikkatyypialaryhmään, on 55 kappaletta. Esimerkkeinä mainittakoon Suo (Maastokohteet), Koski (Vesistökohteet), Kunta, maaseutu (Hallintoalueet ja asutuskohteet) ja Retkeilyalue (Muut kulttuurikohteet).

Toinen tapa jaotella Paikka-kohteet hierarkkisesti on sijaintiin perustuva. Sijaintiin perustuvia hierarkkisia jakoryhmiä ovat seuraavat UML-kaavion arvojoukot:

- Kunta,
- Seutukunta,
- Maakunta,
- Lääni (2009) ja

– Suuralue.

Näistä jakoryhmistä Kunta, Maakunta ja Lääni ovat myös paikkatyyppejä ja näiden arvojoukkojen sisältämille arvoille löytyy vastaava instanssi Paikka-luokasta. Vaikka läänit on lakkautettu vuonna 2010, niin PNR sisältää vuoden 2009 läänejä vastaavat Paikka-instanssit. Sen sijaan tilastollisia alueita (seutukunnat tai vuonna 2011 käytöön otetut uudet suuralueet) vastaavia Paikka-luokan instansseja ei PNR-aineistossa ole.

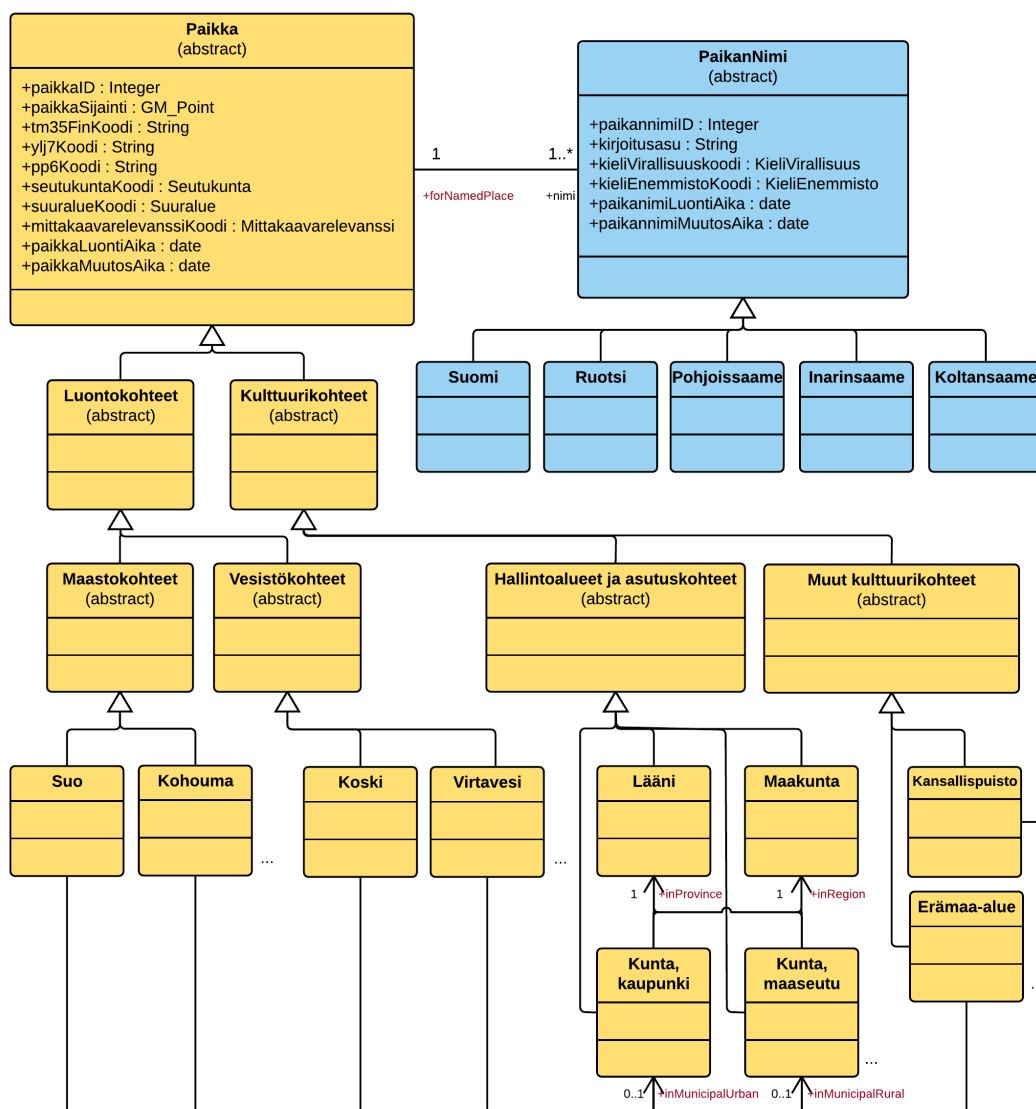
Hierarkioista selkeämpi on paikkatyyppiin perustuva jako, sillä yksi paikkainstanssi edustaa tasan yhtä paikkatyyppiä. Lisäksi, koska GeoSPARQL-ontologia mahdollistaa erilaisten paikkojen välisten topologiasuhteiden ilmaisemisen, tutkimuksessa päädyttiin rakentamaan luokkahierarkiaa paikkatyyppeihin perustuen.

Ennen Tschirnerin ym:n (2011) luomien (kts. luku 4) sääntöjen hyödyntämistä UML-luokkadiagrammin ontologisoimiseksi UML-luokkadiagrammin tietosisältöä muokataan vastaamaan paikkatyyppeihin perustuvaa hierarkkista jaottelua. Näin ollen tehdään Paikka-luokasta abstrakti yläluokka ja luodaan sille abstraktit aliluokat paikkatyyppi- ja paikkatyypialiryhmien mukaisesti (kts. kuva 14). Paikkatyyppien mukaiset luokat, joiden instansseja PNR-aineiston paikkakohteet tulevat linkitettyssä tiedossa olemaan, sijoitetaan paikkatyypialiryhmien aliluokiksi.

Samantyyppinen jako voidaan tehdä Paikannimi-luokan kohteille: jaetaan kohteet eri aliluokkiin nimen kielen perusteella (kts. kuva 14). Tämäkin jako on selkeä, sillä yksi Paikannimi-luokan kohde edustaa tasan yhtä kieltä.

Paikkatyyppien hierarkiaan perustuvan UML-kaavion perusteella voidaan luoda OWL-luokat paikkatyypeistä ja paikkatyypiryhmistä. Kuvassa 15 on määritelty UML-kaavion vasemmanpuoleiset luokat OWL-luokiksi ja OWL-luokille hierarkia. URI-tunnus ontologialle koostuu JHS 193 -suosituksen mukaisesti ”paikkatiedot.fi”-verkkotunnuksesta ja ”/def”-, aineistotunnus- sekä käsitetunnus-polkukomponenteista. Aineistotunnus, joka on annettu kaikille PNR-aineistoon liittyville käsitteille, on ”1001010”. Käsitetunnus PNR-ontologialle on ”pnr”. PNR-ontologian URI-tunnus on siis ”http://paikkatiedot.fi/def/1001010/pnr”, jonka mukaisesta osoitteesta se on myös saatavissa. Ontologiassa määriteltyjen resurssien URI-tunnukset ovat muotoa {PNR-ontologian URI}#{resurssin paikallinen tunnus}. Ontologiassa hyödynnetään ristikkomerkki-URI-tunnuksia (kts. luku 3.2), joten ontologiassa määritetyt resurssit ovat tarjolla kaikki samasta URI-osoitteesta. Koska PNR-ontologiassa GeoSPARQL-ontologia määrittelee ”pnr:NamedPlace”-luokan yläluokan, niin ilmoitetaan GeoSPARQL-ontologia osaksi PNR-ontologiaa ”owl:imports”-käsitteellä.

Kaikille UML-kaavion (kts. kuva 14) luokille luodaan vastaava OWL-luokka Tschirnerin ym:n ensimmäisen säännön (kts. luku 4) mukaisesti. Resurssien tunnusten luomisessa on käytetty aineistossa olevia englanninkielisiä nimiä. ”pnr:NamedPlace” (eli Paikka-luokka) määritellään ”geo:Feature”-luokan aliluokaksi, mikä implikoi, että kyseessä on GeoSPARQL-standardin mukainen paikkatietokohde. Sen sijaan ”pnr:



Kuva 14: Paikannimirekisterin UML-luokkadiagrammi, johon on luotu luokkahierarkia.

GeographicalName” (eli Paikannimi-luokka) on aliluokka suoraan ”owl:thing”-luokalle, joka on kaikkien OWL-luokkien yläluokka. Paikannimi-luokan instanssit eivät siis määrity paikkatietokohteeksi. Näin vältymme turhan toistaiseen tiedon luomiselta.

PNR:n XML-skeematiedostoissa ei ole määritetty erikielisiä nimiä Paikat-tuotteen UML-luokkadiagrammin (kts. kuva 13) mukaisille luokille, arvojoukoille eikä assosiaatio-suhteille. Ontologian ymmärrettävyyden kannalta on kuitenkin hyvä antaa asioille vähintään nimet ja mahdollisuuksien mukaan myös laajemmat kuvaukset. Prototyyppipalvelussa on puuttuvat erikieliset nimet luotu epävirallisesti vapaasti kääntämällä (kts. esimerkiksi kuvan 15 ”pnr:NamedPlace”-luokan predikaatilla ”rdfs:label” ilmaistu suhde).

```

@prefix pnr: <http://paikkatiedot.fi/def/1001010/pnr#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .

<http://paikkatiedot.fi/def/1001010/pnr> a owl:Ontology ;
    owl:imports <http://www.opengis.net/ont/geosparql#> .

pnr:NamedPlace a owl:Class ;
    rdfs:label "Named place"@en,
        "Nimetty paikka"@fi,
        "Ort med namn"@sv ;
    rdfs:subClassOf geo:Feature .

pnr:NaturalFeature a owl:Class ;
    rdfs:label "Natural features"@en,
        "Luontokohteet"@fi,
        "Naturobjekt"@sv ;
    rdfs:subClassOf pnr:NamedPlace .

pnr:TerrainFeature a owl:Class ;
    rdfs:label "Terrain features"@en,
        "Maastokohteet"@fi,
        "Terrängobjekt"@sv ;
    rdfs:subClassOf pnr:NaturalFeature .

pnr:Swamp a owl:Class ;
    rdfs:label "Swamp"@en,
        "Suo"@fi,
        "Myr"@sv ;
    rdfs:subClassOf pnr:TerrainFeature .

```

Kuva 15: PNR-ontologian luokkamäärittelyjä neljälle luokalle.

Koska arvojoukkojen Kunta, Maakunta ja Lääni arvoille löytyy vastaava Paikka-luokan instanssi, niin voidaan Paikka-luokan instanssiin liittyvä kunta, maakunta tai lääni ajatella UML-assosiaationa toiseen Paikka-luokan instanssiin. Näin ollen erillisiä instansseja ei tarvitse luoda kyseisten arvojoukkojen arvoille. Tällä perusteella voidaan luoda siis Tschirnerin ym:n (2011) 4. säännön (kts. luku 4) mukaisesti OWL-kohdeominaisuudet "pnr:inProvince", "pnr:inRegion", "pnr:inMunicipalUrban" ja "pnr:inMunicipalRural" (kts. kuva 16). Maaseutu- ja kaupunkikunnat ovat erotettu, koska ne ovat myös erilliset paikkatyypit PNR-tietomallissa. Nämä kohdeominaisuudet ovat myös topologisia suhteita. Koska GeoSPARQL-ontologia mahdollistaa topologisten suhteiden ilmaisemisen, niin hyödynnetään GeoSPARQL-standardin Simple Features -topologiasanastoa ja määritetään kyseiset kohdeominaisuudet "geo:sfWithin"-

ominaisuuden aliominaisuuksiksi. Vaikka Paikka-kohteiden geometriat ovat pistemäisiä, niin topologiasuhteen luominen voidaan tehdä riippumatta siitä. GeoSPARQL:n topologiasuhteet voidaan yhdistää joko geometriaan ("geo:Geometry") tai kohteeseen ("geo:Feature") (OGC 2012b, s. 7) ja tässä tapauksessa ne yhdistetään kohteeseen.

```

pnr:inMunicipalityRuralArea a owl:ObjectProperty ;
    rdfs:label "Located in municipality"@en,
        "Sijaintikunta"@fi,
        "Belägenhetskommun"@sv ;
    rdfs:domain pnr:NamedPlace ;
    rdfs:range pnr:MunicipalityRuralArea ;
    rdfs:subPropertyOf geo:sfWithin .

pnr:inMunicipalityUrbanArea a owl:ObjectProperty ;
    rdfs:label "Located in city"@en,
        "Sijaintikaupunki"@fi,
        "Belägenhetsstaden"@sv ;
    rdfs:domain pnr:NamedPlace ;
    rdfs:range pnr:MunicipalityUrbanArea ;
    rdfs:subPropertyOf geo:sfWithin .

pnr:inProvince a owl:ObjectProperty ;
    rdfs:label "Located in province"@en,
        "Sijaintilääni"@fi,
        "Belägenhetslän"@sv ;
    rdfs:domain pnr:NamedPlace ;
    rdfs:range pnr:Province ;
    rdfs:subPropertyOf geo:sfWithin .

pnr:inRegion a owl:ObjectProperty ;
    rdfs:label "Located in region"@en,
        "Sijaintimaakunta"@fi,
        "Belägenhetslandskap"@sv ;
    rdfs:domain pnr:NamedPlace ;
    rdfs:range pnr:Region ;
    rdfs:subPropertyOf geo:sfWithin .

```

Kuva 16: Neljän sijaintiin perustuvan OWL-kohdeominaisuuden määrittely. Määritellyt ominaisuudet ovat "geo:sfWithin"-topologiasuhteen aliominaisuuksia.

Paikat-tuotteessa on eksplisiittisesti ilmaistu UML-assosiaatio ainoastaan Paikka-luokasta Paikannimi-luokkaan (kts. kuva 13). Toistaiseen tiedon välttämiseksi ei luoda erillistä geometriaa Paikannimi-luokan instansseille. Sen sijaan luodaan UML-assosiaatio "pnr:forNamedPlace" Paikannimi-luokasta Paikka-luokkaan (kts. kuva 14)

ja määrittellään PNR-ontologiaan vastaava OWL-kohdeominaisuus (kts. kuva 17).

```

pnr:forNamedPlace a owl:ObjectProperty ;
  rdfs:label "For named place"@en,
    "Paikalle"@fi,
    "För ort med namn"@sv ;
  rdfs:domain pnr:GeographicName ;
  rdfs:range pnr:NamedPlace .

```

Kuva 17: OWL-kohdeominaisuuden ”pnr:forNamedPlace” määrittely.

Paikkatyyppien hierarkkisen jaon ja kunta-, maakunta- ja lääni-topologiasuhteiden lisäämisen jälkeen jäljelle jääviä Paikka-luokan ominaisuuksia ovat ”paikkaID”, ”paikkaSijainti”, arvojoukkoihin viittaavat ”seutukuntaKoodi”, ”suuralueKoodi” ja ”mittakaavarelevanssiKoodi” sekä kaikki alkeistyyppeihin viittaavat ominaisuudet kuten ”paikkaLuontiAika”. ”paikkaID”-ominaisuutta hyödynnetään suoraan luotavan paikkatietoinstanssin URI-tunnuksen ”paikallinen tunnus”-osana. Paikka-luokan kuten myös jäljellä olevista Paikannimi-luokan arvojoukoista tehdään Tschirnerin ym:n (2011) 2. säännön mukaisesti OWL-luokkia. Arvojoukkojen arvoista luodaan vastaavasti kyseisten luokkien instansseja. Nämä instanssit tarjotaan JHS 193-suosituksen mukaisesti käsite-URI-tunnuksen kautta. Esimerkiksi ”Mittakaavarelevanssi”-arvojoukosta luotu OWL-luokka saa tunnuksen ”http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi”. Lisäksi luodaan vastaavat OWL-kohdeominaisuudet, joiden sovellusalueet ovat arvojoukoista syntyneet instanssit ja arvoalueet, joko ”pnr:NamedPlace”-luokan instanssit tai ”pnr:GeographicName”-luokan instanssit. Kuvassa 18 on määritelty OWL-kohdeominaisuudet ”pnr:mittakaavarelevanssi” ja ”pnr:kieliVirallisuus”.

```

pnr:mittakaavarelevanssi a owl:ObjectProperty ;
  rdfs:domain pnr:NamedPlace ;
  rdfs:range
    <http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi> .

pnr:kieliVirallisuus a owl:ObjectProperty ;
  rdfs:domain pnr:GeographicName ;
  rdfs:range
    <http://paikkatiedot.fi/def/1001010/KieliVirallisuus> .

```

Kuva 18: OWL-kohdeominaisuuksien ”pnr:mittakaavarelevanssi” ja ”pnr:kieliVirallisuus” määrittely. Nämä määrittelevät implisiittisesti myös arvoalueena esitetyt OWL-luokat. Esimerkiksi ”http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi”.

Paikka- ja Paikannimi-luokan tietotyyppeihin viittaavista ominaisuuksista kuten ”tm35finKoodi” ja ”paikannimiMuutosAika” tehdään OWL-tietotyyppiominaisuuksia

(kts. kuva 19). Tässä yhteydessä olisi mahdollista lisätä tietotyyppiominaisuuksille nimi ja kuvaus esimerkiksi ”rdfs:label”- ja ”dcterms:description”-ominaisuuksilla. Tietotyyppiominaisuuksille voidaan määritellä myös sovellusalue ja arvoalue. Paikka-kohteiden ominaisuuksien arvoalueina ovat mm. ”xsd:integer”-, ”xsd:string”- ja ”xsd:dateTime”-tietotyypit. Eri käytettävien tietotyyppiominaisuuksien arvoalueiden määrittelyjen lisääminen PNR-ontologiaan jatkossa olisi perusteltua.

```

pnr:tm35Fin7Koodi a owl:DatatypeProperty .

pnr:paikannimiMuutosAika a owl:DatatypeProperty .

```

Kuva 19: OWL-tietotyyppiominaisuuksien ”pnr:tm35Fin7Koodi” ja ”pnr:paikkaKorkeus” määrittely.

Arvojoukkojen mukaiset OWL-luokkien instanssit koostuvat kyseisten arvojoukkojen arvoista. Instansseille määritellään URI-tunnukset samoin kuin muulle ontologialle hyödyntäen ristikkomerkki-URI-tunnuksia, jolloin arvojoukon arvoa kuvaavan resurssin kuvaus on saatavilla arvojoukon URI-tunnuksen perusteella. Kuvassa 20 on OWL-luokan Mittakaavarelevanssin sekä yhden esimerkkiarvon määrittely.

```

@prefix Mittakaavarelevanssi:
    <http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi#> .

<http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi>
  a <http://www.w3.org/2002/07/owl#Class> ;
  rdfs:label "Relevance at scale"@en,
    "Mittakaavarelevanssi"@fi,
    "Relevans vid skala"@sv ;
  rdfs:subClassOf
    <http://paikkatiedot.fi/def/1001010/pnr#CodeListPlace> .

Mittakaavarelevanssi:c25000
  a <http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi> ;
  rdfs:label "1:25 000 and larger scales"@en,
    "1:25 000 ja suuremmat mittakaavat"@fi,
    "1:25 000 och större skalor"@sv .

```

Kuva 20: OWL-luokan ”http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi” ja sen instanssin ”Mittakaavarelevanssi:c25000” määrittely.

Arvojoukoista luotujen OWL-luokkien instanssien kuulumisen PNR-ontologiaan ilmaistaan PNR-ontologiassa ”owl:imports”-ominaisuudella, jolloin esimerkiksi ontologian visualisointiin ja selailuun tarvittavat sovellukset osaavat hakea sisällön kyseisen URI-tunnuksen perusteella.

PNR-ontologiassa ei geometriasta mainita mitään muuta kuin, että ”pnr:NamedPlace”-luokka on ”geo:Feature”-luokan aliluokka. Tästä luokkahierarkiasta seuraa, että ”pnr:NamedPlace”-luokan instanssiin voi liittyä ”geo:Geometry”-luokan instanssi ”geo:hasGeometry”-predikaatilla.

”pnr:NamedPlace”-luokan instanssien luomisessa on palvelussa otettu huomioon (toisin kuin alkuperäisessä Paikat-rajapintatuotteessa, joissa esimerkiksi myös lääneillä on kuntakoodi), että maakunnat eivät sijaitse kunnissa tai, että läänit eivät sijaitse maakunnissa. Tämä edesauttaa aineiston ymmärtämistä ja oikeata käyttöä. Suhteet seutukuntiin ja suuralueisiin sen sijaan ovat myös kunnilla, maakunnilla ja lääneillä. Paikka-kohteilla on kohteesta riippumatta viittaus vain yhteen seutukuntaan tai suuralueeseen. Esimerkiksi läänit koostuvat reaaliaikaisessa kuitenkin useasta seutukunnasta. Aineiston metatiedoissa tuleekin selittää, että suhdetieto perustuu ainoastaan Paikka-kohteiden pistegeometrioiden sijaintiin.

5.3 Palvelun toteutus

Paikannimet linkitettynä tietona -palvelun tavoitteena on tarjota PNR-aineisto linkitettynä tietona hyödyntäen WFS-kohdepalvelua. Valmiin rajapintatuotteen hyödyntäminen takaa, että lähtöaineisto on ajantasaista ja, että lähtöaineiston rakenteeseen ei tule yllättäviä muutoksia. Kohdepalvelua hyödynnetään reaaliaikaisesti, jolloin sisältömuutokset tulevat välittömästi näkyviin myös linkitetyssä tiedossa. Toinen tavoite on tarjota aineisto nimenomaan paikkatietokohdekohtaisten URI-tunnusten perusteella. URI-tunnukset pyritään muodostamaan JHS 193 -suosituksen (JUHTA 2015) mukaisesti.

Seuraavassa aliluvussa (luku 5.3.1) esitellään palvelun toteuttamisessa käytetyt ohjelmointikielet, -työkalut ja -kirjastot. Luvussa 5.3.2 kerrotaan luoduista uudelleenohjauskäytänteistä. Luvuissa 5.3.4 ja 5.3.5 esitellään, kuinka HTTP-vastaukset Paikka- ja Paikannimi -kohteiden sekä ontologian tarjoamiseksi luodaan dynaamisesti. Luvussa 5.3.6 esitellään toteutus, jolla koko PNR-aineisto on saavutettavissa PNR-aineistolle annetun URI-tunnuksen perusteella.

5.3.1 Palvelun alusta ja käytetyt työkalut

Paikannimet linkitettynä tietona -prototyypipalvelu on luotu hyödyntäen pelkästään avoimen lähdekoodin ohjelmia, joiden käyttämisestä ei peritä maksua. Palvelun alustana toimii Linux Gentoo virtuaalipalvelin. Palvelu on kehitetty Python ohjelmointikielellä (versio 3.4) hyödyntäen Django Web Framework -sovelluskehystä (kts. <https://djangoproject.com/>). Django-sovelluskehys mahdollistaa toteuttamisessa MVC-arkkitehtuurin (lyhenne sanoista model-view-controller), jossa eritellään järjestelmän tiedon tallentaminen sekä ylläpito, käyttöliittymän ulkoasu sekä tiedon näyttäminen, ja HTTP-pyyntöjen käsittely sekä dynaaminen tiedon muokkaaminen.

Django-sovelluskehyksessä nämä toiminnot on jaettu kolmeen osaan: mallit (engl. models), näkymät (engl. views) ja mallipohjat (engl. templates). Tässä prototyypissä hyödynnetään näkymät-osaa linkitetyn tietosisällön dynaamiseen tuottamiseen ja mallipohjia HTML-muotoisten esitysten luomiseksi. Django sisältää myös prototyypissä hyödynnettävän testipalvelimen dynaamisten web-dokumenttien tarjoamiseksi.

Staatistien tiedostojen tarjoamisesta vastaa Apache HTTP -palvelin (kts. <http://httpd.apache.org/>) Apache HTTP -palvelin mahdollistaa staatistien tiedostojen jakamisen lisäksi ulkoisten ohjelmien ajamisen hyödyntäen moduuleita. Toteutettaessa Paikannimet linkitettynä tietona -palvelu tuotantokäyttöön, täytyy Djangon testipalvelimen sijasta hyödyntää dynaamisten web-dokumenttien tarjoamiseksi esimerkiksi Apache HTTP -palvelimen Python-ohjelmien ajamiseen tarkoitettua `mod_wsgi`-moduulia.

MMML:n WFS-kohdepalvelusta saatavan Paikat -rajapintatuotteen GML-muotoisen tiedon käsittelyssä hyödynnetään Pythonin omaa The ElementTree XML API -kirjastoa (kts. <https://docs.python.org/3.4/library/xml.etree.elementtree>). Kirjastossa on määritelty Element-luokka, jonka instanssille voidaan määritellä mm. nimi, sisältö, ominaisuustiedot ja lapsielementit. Kirjastossa on metodi, joka osaa lukea XML-tiedoston ja luoda siitä Element-luokan instansseista koostuvan ElementTree-luokan instanssin.

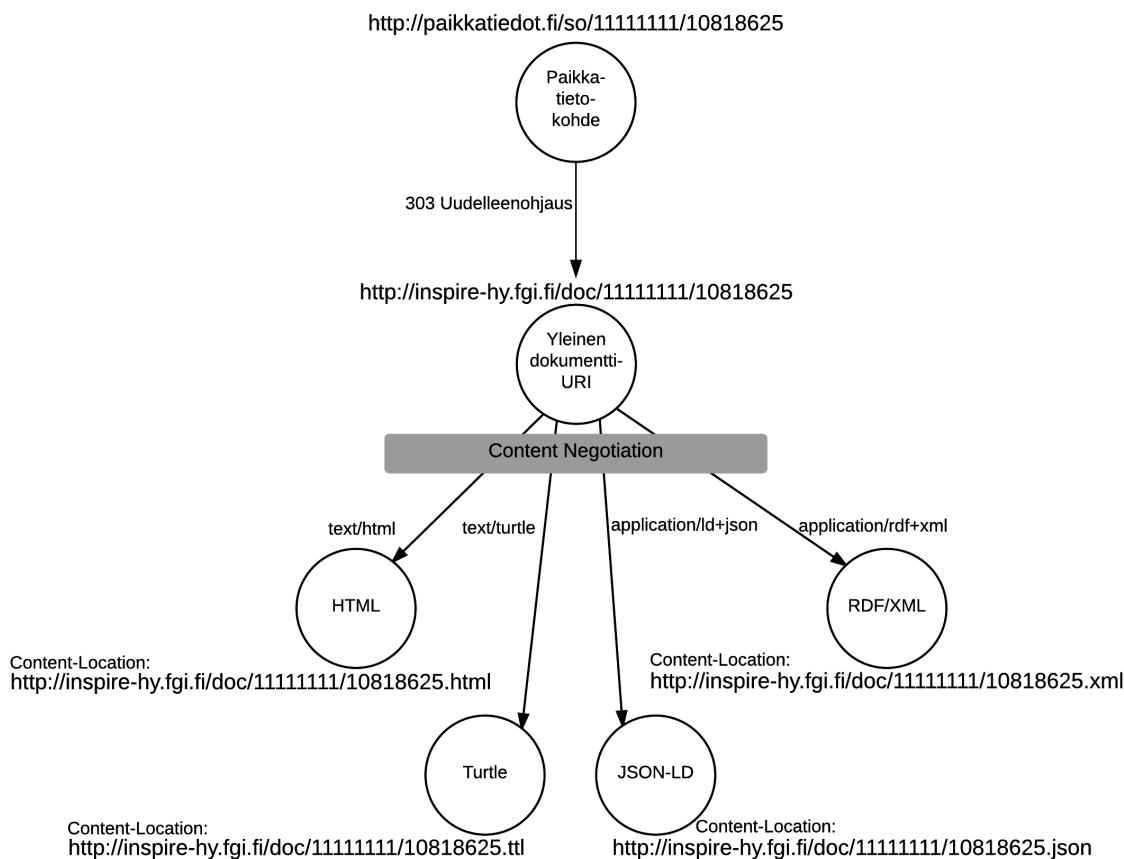
RDF-tiedon käsittelyssä hyödynnetään Pythonille luotua RDFLib-kirjastoa (kts. <https://github.com/RDFLib/rdfli>). Kirjastossa on määritelty Graph-luokka, joka mahdollistaa RDF-tiedon monipuolisen muokkaamisen ja käsittelyn. Kirjasto sisältää mm. metodit graafien sarjallistamiseksi eri muotoihin, kuten RDF/XML- tai Turtle-muotoon.

HTML-muotoisten esitysten ulkoasun luomisessa hyödynnetään Bootstrap-kirjastoa (kts. <http://getbootstrap.com/>). Paikka-kohteille toteutettu karttanäkymä on toteutettu OpenLayers 3 - JavaScript-kirjaston (kts. <http://www.openlayers.org>) avulla.

5.3.2 URI-tunnusten ja uudelleenohjausten määrittäminen

PRN linkitettynä tietona -prototyypipalvelun verkkotunnus on ”inspire-hy.fgi.fi”. JHS 193 -suositus ei varsinaisesti määrittele paikkatietoaineistolle omaa URI-tunnusta. Tässä työssä aineiston URI-tunnus on luotu hierarkkisesti suoraan paikkatietokohdeiden tunnuksen mukaisesti jättämällä pois paikkatietokohteen paikallinen tunnus. Näin ollen PNR-aineistolle annettava Paikka-kohteiden aineisto-URI-tunnus on ”<http://paikkatiedot.fi/so/11111111/>”. Aineisto-URI-tunnus ja kaikki muut HTTP-pyynnöt sillä alkaviin URI-tunnuksiin on uudelleenohjattu ”<http://inspire-hy.fgi.fi/doc/11111111/>”-alkuisiin osoitteisiin (kts. kuva 21). Paikannimi-kohteiden aineisto-URI-tunnus on ”<http://paikkatiedot.fi/so/22222222/>”, jolle on tehty vastaavat uudelleenohjaukset. Vastaavasti kaikki ”<http://paikkatiedot.fi/def/1001010/>”-alkuiset käsite-URI-tunnukset, kuten ontologian ja arvojoukkojen tarjoamiseksi tarkoitettut,

on asetettu ohjautumaan ”http://inspire-hy.fgi.fi/ont/”-alkuisiin osoitteisiin. Prototyyppisovellus hyödyntää Django-sovelluskehityksen testipalvelinta, joka on asetettu toimimaan portissa 8000. Jotta URI-tunnuksiin ei tarvitsisi lisätä ylimääräisiä portti-määrittelyjä, on Apache HTTP -palvelimeen luotu alias ohjaamaan ”/doc”- ja ”/ont”-alkuisiin polkuosioihin kohdistuvat HTTP-pyyntö kyseiseen porttiin.



Kuva 21: Palvelun uudelleenohjauksen periaate.

HTTP-pyyntö, jotka päätyvät Django-palvelimen käsiteltäväksi, ohjataan Views-moduuliin oikean Python-funktion käsiteltäväksi. Ohjaukset luodaan Django:n Urls-moduuliin regex-määrittelyiden (engl. regular expression) avulla. Views-moduulissa määrittelyssä funktiossa luodaan sisältö palautettavaan HTTP-vastaukseen.

Palautettavan HTTP-vastauksen sisältöön vaikuttaa URI-tunnuksen sisältämät tiedot, kuten paikallinen tunnus -osio. Views-moduulin funktioissa voidaan myös tutkia HTTP-pyyntönsä otsaketietoja tai pyyntöön liittyviä GET-parametreja. Content Negotiation perustuu HTTP-pyyntönsä Accept-otsakkeeseen. Content Negotiation-toteutuksessa otetaan huomioon HTTP-pyyntönsä Accept-otsakkeeseen lisätty tietoformaattitunnukset niiden esiintymisjärjestyksessä. Näin tieto voidaan palauttaa pyydetyssä tietoformaattissa. Tunnuksille on mahdollista lisätä myös laatu-arvo (engl. quality value, W3C 2008) parhaiten sopivan palautusmuodon tarjoamiseksi, mutta

tätä ominaisuutta ei ole prototyypissä hyödynnetty. HTTP-vastaukseen lisätään Content-Location -otsake vastauksen muodon mukaisesti (kts. kuva 21).

5.3.3 Palvelun toimintaperiaate

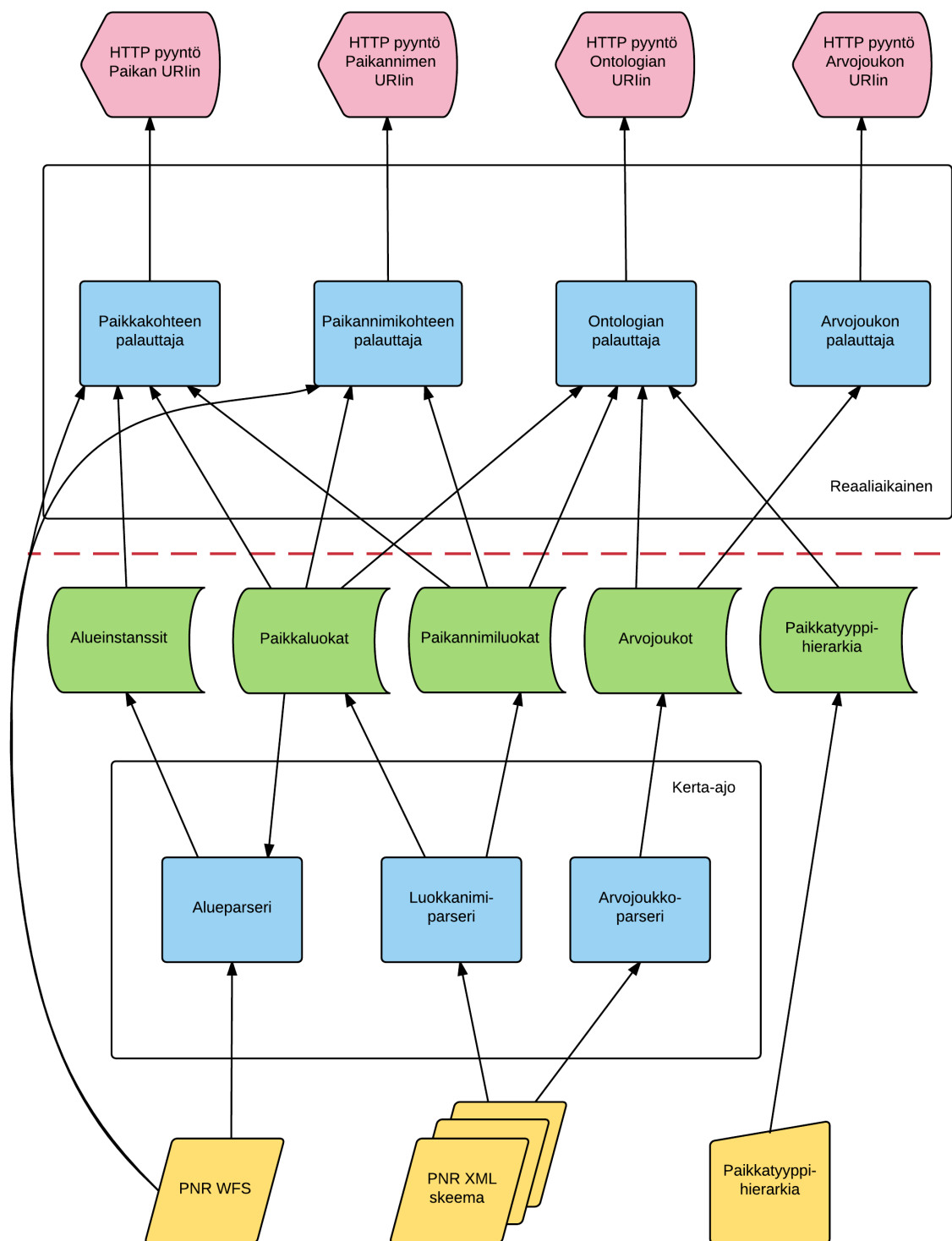
Paikannimet linkitettyinä tietona -palvelun toiminnot tapahtuvat kaksissa eriaikaisissa prosesseissa: kerta-ajoprosesseissa ja reaaliaikaisissa prosesseissa. Kerta-ajona suoritettavat prosessit on toteutettu Python-ohjelmointikielellä ja ne ajetaan ilman Django -web-sovelluskehystä paikallisesti palvelimelta. Kerta-ajoprosessien avulla voidaan mm. vähentää reaaliaikaisissa prosesseissa tarvittavien HTTP-pyyntöjen määrää ja näin huomattavasti nopeuttaa palvelun toimintaa. Kuvassa 22 on esitelty palvelun toimintaperiaate ja kerta-ajo- sekä reaaliaikainen prosessit, joiden avulla voidaan luoda ja tarjota paikka- ja paikannimikohteet sekä ontologia ja siihen liittyvät arvojoukot.

Kerta-ajoprosesseja (kts. laatikko kuvan 22 alaosassa) ovat Alueparseri, Luokkanimiparseri ja Arvojoukkoparseri. Näiden prosessien tehtävänä on hakea PNR:n XML-skeematiedostojen osana olevat Paikkatyyppi-, Paikkatyyppiryhmä-, Kieliarvojoukko sekä muut arvojoukot ja luoda niiden tietosisältöä vastaavat ontologian mukaiset paikka- ja paikannimiluokat sekä muita arvojoukkoja vastaavat OWL-luokat ja niiden instanssit. Luoduille luokille ja arvojoukkojen instansseille annetaan viitetiedoksi XML-skeemassa oleva koodi ”pnr:hasCode”-predikaatilla. Alueparseri hakee lisäksi PNR:n WFS-kohdepalvelusta kaikki kunta-, maakunta- ja lääni-instanssit ja niille vastaavat kunta-, maakunta- tai läänikoodit, jotka tallennetaan käyttäen ”pnr:hasCode”-predikaattia. Kaikki kerta-ajona luodut tiedot tallennetaan linkitettyinä tietona Turtle-muodossa tiedostoihin (kts. vihreät säiliöt kuvassa 22). XML-skeematiedostoista löytyvät suomen-, ruotsin- ja englanninkieliset nimet paikkatyypeille, paikannimien kielille ja muiden arvojoukkojen kohteille tallennetaan luoduille OWL-luokille ja instansseille käyttäen ”rdfs:label”-predikaattia.

Reaaliaikaiset prosessit (kts. laatikko kuvan 22 yläosassa) suoritetaan aina vastaavan HTTP-kutsun yhteydessä. Prosessi hyödyntää kerta-ajona luotua linkitettyä tietoa sekä PNR:n WFS-kohdepalvelua ajantasaisen tiedon tarjoamiseksi. Seuraavissa aliluissa on tarkemmin esitelty ontologian/arvojoukkojen, paikka/paikannimi-kohteiden sekä aineiston/osa-aineistojen (huom. viimeistä ei ole piirretty kuvaan 22) tarjoaminen dynaamisesti.

5.3.4 Ontologian tarjoaminen

Ontologia luodaan dynaamisesti hyödyntäen kerta-ajoprosesseissa tallennettuja tietoja. Python-funktiossa ontologian rakentamiseksi luodaan graafi ja määritellään käytettävät nimiavaruudet. Tallennetut kerta-ajoprosesseissa luodut Turtle-tiedostot tuodaan RDFLib-kirjaston avulla RDF-graafeiksi. Graafeista on yksinkertaista hakea ja yhdistellä ontologia-graafiin tarvittavia tietoja RDFLib-kirjaston metodien



Kuva 22: Palvelun toimintaperiaate. Nuolet kuvastavat pääasiallista informaation kulun suuntaa.

avulla. Ontologia-graafiin lisätään luvussa 5.2 esitelty sisältö paikkatyyppihierarkioineen. Ontologiaan liittyvien arvojoukkojen luominen (esim. HTTP-pyyntö osoit-

teeseen <http://paikkatiedot.fi/def/1001010/Mittakaavarelevanssi>) toteutetaan vastaavasti. Paikkatyypeille, paikkatyyppiryhmille sekä arvojoukkojen instansseille löytyy PNR:n XML-skeemoista nimet kolmella eri kielellä. Sen sijaan arvojoukoille ja luokille ei löydy kuin suomenkielinen otsikko. Näille, kuten myös ontologisoinnissa luoduille assosiaatiosuhteita ilmaiseville predikaateille on Ontologian palauttaja -prosessin ohjelmakoodiin lisätty nimet kolmella eri kielellä.

Ontologian tai siihen liittyvien arvojoukkojen sisältö luodaan RDF-graafiksi aina HTTP-pyyntöön yhteydessä. Riippuen HTTP-pyyntöön lähettäjän toiveista tiedon esitysmuodon suhteen graafi sarjallistetaan joko RDF/XML-, Turtle- tai JSON-LD-muotoon ja palautetaan HTTP-vastauksena. Jos käyttäjä pyytää HTTP-vastauksen HTML-muodossa, niin palvelu palauttaa dynaamisesti luodun web-sivun, josta on linkitys kahteen eri ontologia-katselupalveluun Parrot (kts. <http://ontorule-project.eu/parrot/parrot>) ja WebProtege (kts. <http://webprotege.stanford.edu/>). Parrot-palvelu luo dynaamisesti katseltavan version koko ontologiasta, kun taas WebProtege-palveluun on PNR-ontologia lisätty erikseen.

5.3.5 Paikka- ja paikannimikohteiden tarjoaminen

Paikka- tai paikannimikohteen URI-tunnukseen tehdyn HTTP-pyyntöön perusteella pyyntö ohjataan oikealle Python-funktiolle. URI-tunnuksen paikallinen osa eli PNR:n paikkaID tai paikannimiID tallentuu funktion parametriksi. Vastauksen luomisessa hyödynnetään PNR:n WFS-kohdepalvelua. WFS-kyselyssä haetaan PNR:n Paikka- tai Paikannimi-kohdetta WFS-kohdepalvelun "FeatureID"-kyselyparametria hyödyntäen. Jos URI-tunnusta vastaava kohde löytyy, niin WFS-kohdepalvelusta saatu GML-tieto muunnetaan suunnitellun ontologian mukaisesti RDF-tiedoksi. Jos URI-tunnuksen paikallinen osa ei vastaa mitään PNR:n kohdetta, niin palautetaan HTTP-vastaus statuskoodilla "404 NOT FOUND".

RDF-graafin luomisessa GML-tiedosta hyödynnetään kerta-ajoprosesseissa tallennettuja tietoja, kuten kunta, maakunta ja lääni-instansseja. GML-tiedossa olevien koodien perusteella voidaan paikkatietokohdetta kuvaavaan graafiin liittää ontologian mukainen tieto. Kuvassa 23 on dynaamisesti luotu Turtle-muotoinen RDF-esitys paikkatietokohteesta. Kyseinen Paikka-kohde viittaa kahteen eri Paikannimi-kohteeseen. Paikannimi-kohteista on esitetty palautetussa graafissa ainoastaan nimitieto. Lisätiedon saamiseksi on lähetettävä HTTP-pyyntö kyseiseen Paikannimi-kohteen URI-osoitteeseen. Paikannimi-kohteen RDF-muotoinen esitys (kts. kuva 24) luodaan vastaavasti hakemalla kohteen tiedot WFS-kohdepalvelusta ja hyödyntäen kerta-ajoprosesseissa tallennettuja tietoja.

Käyttäjien tarpeet määrittävät osin sen, mitä kaikkea resurssiin liittyvää tietoa tulisi tarjota yhden URI-tunnuksen kautta. Tässä palvelussa on käytetty ratkaisua, jossa itse URI-tunnuksen mukaisesta resurssista tarjotaan kaikki olemassa oleva tieto. Sen lisäksi tarjotaan nimitietoja muista resursseista, joihin tästä resurssista


```

<http://paikkatiedot.fi/so/11111111/10342061> a pnr:Elevation ;
  rdfs:label "Kotkavuori" ,
    "Kotkavuori"@fi ,
    "Örnberget"@sv ;
  pnr:hasGeographicName
    <http://paikkatiedot.fi/so/22222222/40342061> ,
    <http://paikkatiedot.fi/so/22222222/40903882> ;
  pnr:inMunicipalityUrbanArea
    <http://paikkatiedot.fi/so/11111111/10342733> ;
  pnr:inProvince <http://paikkatiedot.fi/so/11111111/10899564> ;
  pnr:inRegion <http://paikkatiedot.fi/so/11111111/10899566> ;
  pnr:mittakaavarelevanssi <http://paikkatiedot.fi/def/1001010/
    Mittakaavarelevanssi#c25000> ;

  pnr:paikkaKorkeus 24 ;
  pnr:paikkaLuontiAika "2008-12-06T00:00:00.000"^^xsd:datetime ;
  pnr:paikkaMuutosAika "2014-02-27T00:00:00.000"^^xsd:datetime ;
  pnr:pp6Koodi "1904A1" ;
  pnr:seutukunta
    <http://paikkatiedot.fi/def/1001010/Seutukunta#c011> ;
  pnr:suuralue <http://paikkatiedot.fi/def/1001010/Suuralue#c1> ;
  pnr:tm35Fin7Koodi "L4133A2" ;
  pnr:ylj7Koodi "203403C" ;
  void:inDataset <http://paikkatiedot.fi/so/11111111:091335/> ;
  geo:hasGeometry
    [ geo:asWKT " <http://www.opengis.net/def/crs/EPSG/0/3067>
      Point(382533.781 6671382.603)"^^geo:WKTLiteral ] ;
  rdfs:isDefinedBy
    <http://inspire-hy.fgi.fi/doc/11111111/10342061> .

<http://paikkatiedot.fi/so/11111111:091335/> dcterms:title
  "Places: Helsinki, Elevation"@en,
  "Paikat: Helsinki, Kohouma"@fi,
  "Platserna: Helsingfors, Upphöjning"@sv .

<http://paikkatiedot.fi/so/22222222/40342061> a pnr:FinnishName ;
  rdfs:label "Kotkavuori"@fi .
<http://paikkatiedot.fi/so/22222222/40903882> a pnr:SwedishName ;
  rdfs:label "Örnberget"@sv .

```

Kuva 23: Paikka-kohde esitettynä RDF-tietona Turtle-sarjallistamismuodossa.

viitataan. Nimet helpottavat graafin tulkintaa ilman sovellusta, ja esimerkiksi ihmisluettavan HTML-sivun luominen kyseisestä graafista vaatii näin vähemmän erillisiä HTTP-pyyntöjä. Mitä enemmän tietoa saadaan yhden URI-tunnuksen kautta, sitä

```

<http://paikkatiedot.fi/so/22222222/40342061> a pnr:FinnishName ;
    rdfs:label "Kotkavuori",
        "Kotkavuori"@fi ;
    pnr:forNamedPlace <http://paikkatiedot.fi/so/11111111/10342061> ;
    pnr:hasParallelName
        <http://paikkatiedot.fi/so/22222222/40903882> ;
    pnr:kieliEnemmisto
        <http://paikkatiedot.fi/def/1001010/KieliEnemmisto#c1> ;
    pnr:kieliVirallisuus
        <http://paikkatiedot.fi/def/1001010/KieliVirallisuus#c1> ;
    pnr:kirjoitusasu "Kotkavuori" ;
    pnr:paikannimiLahde
        <http://paikkatiedot.fi/def/1001010/PaikannimiLahde#c1> ;
    pnr:paikannimiLuontiAika "2008-12-06T00:00:00.000"^^xsd:datetime ;
    pnr:paikannimiMuutosAika "2008-12-06T00:00:00.000"^^xsd:datetime ;
    pnr:paikannimiStatus
        <http://paikkatiedot.fi/def/1001010/PaikannimiStatus#c5> ;
    rdfs:isDefinedBy
        <http://inspire-hy.fgi.fi/doc/22222222/10342061> .

<http://paikkatiedot.fi/so/11111111/10342061> a pnr:Elevation ;
    rdfs:label "Kotkavuori",
        "Kotkavuori"@fi,
        "Örnberget"@sv .

<http://paikkatiedot.fi/so/22222222/40903882> a pnr:SwedishName ;
    rdfs:label "Örnberget"@sv .

```

Kuva 24: Paikkanimi-kohde esitettynä RDF-tietona Turtle-sarjallistamismuodossa.

vähemmän tarvitsee tehdä HTTP-pyyntöjä. Toisaalta toisteista ja ylimääräistä tietoa tulisi välttää, jotta tiedonsiirto olisi mahdollisimman tehokasta.

Vastaavasti kuin ontologian ja ontologiaan liittyvien arvojoukkojen tapauksessa, luodaan Paikka- ja Paikannimi-kohteista ensin RDF-graafi RDFLib-kirjaston avulla. Kohde sarjallistetaan HTTP-pyyntöön toivomaan tietoformattiin ja palautetaan vastauksena. Sarjallistaminen onnistuu yksinkertaisesti RDFLib-kirjaston metodeilla. HTML-muotoinen esitys Paikka- ja Paikannimikohteista luodaan dynaamisesti hyödyntämällä HTML-taulukkoita (kts. kuva 25). Taulukoihin tulevat riveittäin aina resurssia kuvaavan predikaatin nimi ja kyseinen arvo (eli objekti). Lisäksi predikaatit ja mahdollisesti myös arvot, jos nämä ovat itse resursseja, toimivat linkkeinä omiin URI-osoitteisiinsa. Paikka-kohteille näytetään myös pieni OpenLayers-kirjastolla toteutettu karttaikkuna, jossa paikan sijainti on merkitty pisteellä. Taustakarttana hyödynnetään MML:n taustakarttaa WMS-palvelun välityksellä.

MML Nimistö

<http://paikkatiedot.fi/so/11111111/10342061>

Nimetty paikka

Kotkavuori Kohouma



Paikan nimi

ruotsi: Örnberget

suomi: Kotkavuori

Paikkatyyppi

Kohouma

Maastokohteet

Luontokohteet

Sijainti

Helsinki

Uusimaa

Etelä-Suomi

Koordinaatit

ETRS-TM35FIN: 382534, 6671383

WGS84: 24.8834, 60.16242

Lisätietoja

Suuralue: Helsinki-Uusimaa

Seutukunta: Helsinki

Mittakaavarelevanssi: 1:25 000 ja suuremmat mittakaavat

in dataset: Paikat: Helsinki, Kohouma

Korkeus: 24

TM35FIN7 koodi: L4133A2

YLJ7 koodi: 203403C

PP6 koodi: 19O4A1

LuontiAika: 2008-12-06T00:00:00

Päivitysaika: 2014-02-27T00:00:00

Saatavilla formateissa (Content Negotiation): [RDF/XML](#) [Turtle](#) [JSON-LD](#) [Schema.org](#)[suomi](#) [ruotsi](#) [englanti](#)

Kuva 25: Paikka-kohteesta dynaamisesti luotu HTML-muotoinen esitys.

Hakukoneyhtiöiden tukema schema.org-sanasto mahdollistaa rakenteisen tiedon tulokinnan ja hyödyntämisen hakukoneissa. Google (2015) ohjeistaa lisäämään schema.org -sanaston mukaisen rakenteisen tiedon JSON-LD-muodossa HTML-sivulla "script"-elementin sisään. HTML-sivun dynaamisen luomisen yhteydessä luodaan siis schema.org -sanaston mukainen "Place"-luokan instanssi (kts. kuva 26), joka sisältää kohteen nimen, tiedon siitä, missä toisessa kohteessa tämä kohde sijaitsee, kohteen kuvauksen ja koordinaattipisteen WGS84 - maantieteellisessä koordinaatistossa.

```

{
  "@context": "http://schema.org",
  "@id": "http://paikkatiedot.fi/so/11111111/10342061",
  "@type": "Place",
  "alternateName": "Örnberget@sv, Kotkavuori@fi",
  "containedIn": {
    "@id": "http://paikkatiedot.fi/so/11111111/10342733",
    "@type": "Place",
    "name": "Helsinki",
    "url": "http://paikkatiedot.fi/so/11111111/10342733"
  },
  "description":
    "Place type: Elevation in NLS Finland Place Name Register",
  "geo": {
    "@type": "GeoCoordinates",
    "elevation": "24",
    "latitude": 60.16242,
    "longitude": 24.8834
  },
  "name": "Kotkavuori"
}

```

Kuva 26: Paikkanimi-kohde esitettynä schema.org -sanaston ”Place”-luokan kohteena JSON-LD-muodossa.

5.3.6 Aineistolinkitysten toteuttaminen

Berners-Leen (2006) linkitetyn data neljännen periaatteen mukaan tulee resursseille lisätä mahdollisuuksien mukaan linkkejä toisiin kohteisiin. Tähän mennessä Paikka-kohteet linkittyvät vain osittain toisiinsa ”geo:sfWithIn”-predikaatilla. Tämän takia otetaan käyttöön lisäksi VoID-sanasto, jota voidaan käyttää resursseista koostuvien aineisto-resurssien kuvailuun. Paikkanimet linkitettynä tietona -palvelun aineistotunnus ”http://paikkatiedot.fi/so/11111111/” ohjaa dokumenttiin ”http://inspire-hy.fgi.fi/doc/11111111/”, joka kuvaa aineistoa. Aineiston RDF-muotoinen esitys (kts. kuva 27) sisältää mm. aineisto-resurssin tyypin, otsikon, käyttöoikeuslisenssin tiedot, aineiston tarjoajan, saatavilla olevat formaatit sekä aineiston aiheen.

Paikka-kohteita on PNR:n WFS-kohdepalvelussa 797659 (haku 1.10.2015), joten käsiteltävyyden vuoksi kaikkien kohteiden tarjoaminen yhtenä listana ei ole järkevää. Toinen rajoittava tekijä on kyseisen kohdepalvelun palauttamien kohteiden maksimimäärä, joka on 10000. ”void:subset”-predikaatilla on mahdollista määritellä aineistolle osa-aineistoja. Aineiston jakaminen osa-aineistoihin mahdollistaa isonkin aineiston kaikkien kohteiden URI-tunnusten tarjoamisen linkitysten avulla. Aineisto on jaettu kahdella eri jakoperusteella, kunnittain ja paikkatyypeittäin, hierarkkisesti kahteen

```

<http://paikkatiedot.fi/so/11111111/> a void:Dataset ;
  dcterms:contributor [ a foaf:Organization ;
    rdfs:label "National Land Survey of Finland"@en,
    "Maanmittauslaitos"@fi,
    "Lantmäterieverket"@sv ] ;
  dcterms:description
    "Paikannimirekisterin data linkitettyinä
      tietona kohdepalvelun kautta"@fi,
  dcterms:license <http://creativecommons.org/licenses/by/4.0/> ;
  dcterms:subject <http://dbpedia.org/resource/Place_(geography)> ;
  dcterms:title "Places of Place Name Register"@en,
    "Paikannimirekisterin paikat"@fi,
    "Ortnamns registerns orter"@sv ;
  void:feature <http://www.w3.org/ns/formats/JSON-LD>,
    <http://www.w3.org/ns/formats/RDF_XML>,
    <http://www.w3.org/ns/formats/Turtle> ;
  void:subset
    <http://paikkatiedot.fi/so/11111111:005/>,
    <http://paikkatiedot.fi/so/11111111:992/> ;
  ...
  rdfs:isDefinedBy <http://inspire-hy.fgi.fi/doc/11111111/> .

<http://paikkatiedot.fi/so/11111111:005/> a void:Dataset ;
  dcterms:title "Places: Alajärvi"@en,
    "Paikat: Alajärvi"@fi,
    "Orterna: Alajärvi"@sv .
<http://paikkatiedot.fi/so/11111111:992/> a void:Dataset ;
  dcterms:title "Places: Äänekoski"@en,
    "Paikat: Äänekoski"@fi,
    "Orterna: Äänekoski"@sv .
...

```

Kuva 27: PNR Paikka-aineiston RDF-muotoinen esitys. (Esimerkkiin on sisällytetty vain kahden osa-aineiston tiedot).

osa-aineistotasoon. Näistä ensimmäisen tason osa-aineistot (eli kunnat) sisältävät vain osa-aineistoja ja toisen tason osa-aineistot (eli paikkatyypit kunnittain) sisältävät itse paikkatietokohteet. Tällaisella jaolla paikkojen määrä ei kasva kohtuuttoman suureksi yhdessä toisen tason osa-aineistossa ja pärjätään yhdellä alle 10000 kohdetta palauttavalla WFS-kyselyllä.

Paikkatietoaineiston ollessa kyseessä on sijainti, tässä tapauksessa kunta, valittu ensimmäisen tason jakoryhmäksi. JHS 193 -suositus ei määrittele käsitettä osa-aineisto. Osa-aineistojen tunnuksen luonti on ratkaistu lisäämällä luotuun aineisto-

URI-tunnuksen aineistotunnusosioon kaksoispiteen jälkeen osa-aineiston mukainen kuntakoodi. Esimerkiksi ensimmäisen tason osa-aineisto nimeltä ”Paikat: Äänekoski” saa tunnuksen ”http://paikkatiedot.fi/so/11111111:992/”. Tätä URI-tunnusta vastaava dokumentti-URI ”http://inspire-hy.fgi.fi/doc/11111111:992/” palauttaa osa-aineiston kuvauksen (kts. kuva 28). Kuvauksessa määritellään mm. mihin aineistoon kyseinen osa-aineisto kuuluu ja miten tämä aineisto jakautuu uusiin toisen tason osa-aineistoihin.

```
<http://paikkatiedot.fi/so/11111111:992/> a void:Dataset ;
    dcterms:title "Places, Äänekoski"@en,
        "Paikat, Äänekoski"@fi,
        "Orterna, Äänekoski"@sv ;
    void:inDataset <http://paikkatiedot.fi/so/11111111/> ;
    void:subset <http://paikkatiedot.fi/so/11111111:992335/>,
        <http://paikkatiedot.fi/so/11111111:992435/>,
        <http://paikkatiedot.fi/so/11111111:992325/> ;
    ...
    rdfs:isDefinedBy <http://inspire-hy.fgi.fi/doc/11111111:992/> .

<http://paikkatiedot.fi/so/11111111:992435/> a void:Dataset ;
    dcterms:title "Places: Äänekoski, Rapids"@en,
        "Paikat: Äänekoski, Koski"@fi,
        "Orterna: Äänekoski, Fors"@sv
<http://paikkatiedot.fi/so/11111111:992325/> a void:Dataset ;
    dcterms:title "Places: Äänekoski, Forest"@en,
        "Paikat: Äänekoski, Metsäalue"@fi,
        "Orterna: Äänekoski, Skogsområde"@sv .
...
```

Kuva 28: Ensimmäisen tason osa-aineiston RDF-muotoinen esitys. (Esimerkkiin on sisällytetty vain kahden osa-aineiston tiedot).

Toisen tason osa-aineistot jaetaan paikkatyypeittäin. Aineisto-URI-tunnus muodostetaan lisäämällä kuntakoodin perään paikkatyyppikoodi. Python-funktio, joka käsittelee toisen tason aineisto-URI-tunnukseen tehdyn pyynnön saa parametrikseen kunta- ja paikkatyyppikoodin. Näiden tietojen perusteella, hyödyntäen Filter Encoding-rajausehtoja, haetaan PNR:n WFS-kohdepalvelusta kaikki kyseisen kunta- ja paikkatyyppikoodin sisältämät Paikka-kohteet. Esimerkiksi osa-aineisto nimeltä ”Paikat: Äänekoski, Metsäalue” saa tunnuksen ”http://paikkatiedot.fi/so/11111111:992325/” ja tätä URI-tunnusta vastaava dokumentti-URI palauttaa toisen tason osa-aineiston, joka sisältää Äänekoskella sijaitsevat metsäalueet (kts. kuva 29). Osa-aineistoon kuuluva paikkatietokohde on esimerkiksi ”Kettukorpi”. VoID-sanasto ei sisällä tapaa ilmaista, mitkä kohteet kuuluvat ”void:Dataset”-luokan instanssiin. Sen sijaan on olemassa predikaatti ”void:inDataset”, jolla toisen suuntainen suhde voidaan ilmais-

ta. Molempien suuntaisen linkityksen vuoksi tarjotaan osa-aineiston URI-osoitteen palauttamassa graafissa myös siihen sisältyvät kohteet erillisinä resursseina. Resursseille lisätään nimitieto ja ”void:inDataset”-predikaatilla tieto, että kohteet kuuluvat kyseiseen osa-aineistoon (kts. kuva 29).

```
<http://paikkatiedot.fi/so/11111111:992325/> a void:Dataset ;
    dcterms:title "Places: Äänekoski, Forest"@en,
        "Paikat: Äänekoski, Metsäalue"@fi,
        "Orterna: Äänekoski, Skogsområde"@sv ;
    void:inDataset <http://paikkatiedot.fi/so/11111111:992/> ;
    rdfs:isDefinedBy <http://inspire-hy.fgi.fi/doc/11111111/992325> .

<http://paikkatiedot.fi/so/11111111/10353809> a pnr:Forest ;
    rdfs:label "Kettukorpi"@fi ;
    void:inDataset <http://paikkatiedot.fi/so/11111111:992325/> .

...
```

Kuva 29: Toisen tason osa-aineiston RDF-muotoinen esitys. (Esimerkkiin on sisällytetty vain yhden kohteen tiedot).

Paikannimi-kohteille ei vastaavaa aineistolinkitystä ole luotu. Paikannimet linkittyvät joka tapauksessa samaan Paikka-aineistoon Paikka-instanssien kautta ”pnr:hasName”-predikaatilla ja sitä vastaavaa käänteistä suhdetta ilmaisevalla ”pnr:forNamedPlace”-predikaatilla.

HTML-muodossa aineistot ja osa-aineistot tarjoavat käyttäjälle mahdollisuuden selata aineistoa linkkien avulla. RDF-tiedossa on kohteille, jotka kuuluvat toisen tason osa-aineistoon, lisätty nimi, jolloin HTML-muodossa voidaan näyttää linkin yhteydessä nimi ja järjestää kohteet aakkosjärjestykseen. Näin muutamankin tuhannen kohteen listasta (esimerkiksi aineisto ”http://paikkatiedot.fi/so/11111111:740570/” nimeltä ”Paikat: Savonlinna, Talo”) on selainsovelluksella haettavissa tietty kohde.

5.4 Palvelun hyöty ja käyttötapauksia

Työssä suunnitellun linkitetyn tiedon palvelun on tarkoitus parantaa PNR-aineiston saavutettavuutta ja käytettävyyttä. Webin yleisten tekniikoiden käyttäminen paikkatietoaineiston tarjoamisessa antaa mahdollisuuden lukuisten erilaisten sovellusten ja tekniikoiden käyttämiseen myös aineiston hyödyntämisessä. Paikkatietokohteista, aineistosta ja osa-aineistoista tarjottavat HTML-muotoiset esitykset mahdollistavat aineistoon tutustumisen pelkkän web-selaimen avulla. Aineiston metatiedot voidaan laittaa saataville suoraan aineisto-URI-osoitteeseen, josta pääsee myös selailemaan aineiston sisältöä aineistolinkityksen avulla. Palvelussa käytetyt käsittehierarkiat ja asioiden välisiä suhteita kuvaavat käsitteet ovat saatavilla selityksineen ontologian

muodossa yksinkertaisten linkkien välityksellä. Tämä kaikki voi auttaa käyttäjää saamaan kuvan aineiston sisällöstä, tarkoituksesta ja laadusta, ja mahdollistaa näin aineiston järkevän hyödyntämisen.

Paikoille annettavat yksilöivät ja pysyvät URI-tunnukset mahdollistavat PNR-aineiston paikkojen linkittämisen toisiin aineistoihin. Monissa aineistoissa, kuten esimerkiksi tutkimusartikkeleissa tai tietorekistereissä, on viittauksia paikkoihin. Käyttämällä näissä aineistoissa paikkojen nimistä viittauksia PNR:n paikkoihin mahdollistetaan paikannimien yksikäsitteisyys ja ajantasaisuus. Koska viittauksessa käytetään URI-tunnusta, saadaan paikasta myös varsin kätevästi lisätietoa URI-tunnuksen mukaisesta osoitteesta. Kuvassa 30 on esimerkki uutisartikkeliin lisättävästä viittauksesta Paikka-kohteeseen.

Puolustusvoimat aloitti tonttien pakkolunastuksen

Lunastettavat alueet ovat ampuma-alueen vieressä Taipalsaarella

KOTIMAA 6.7.2015 21:01 Päivitetty: 6.7.2015 21:49

Outi Salovaara HELSINGIN SANOMAT

Sarviniemen alueen pakkolunastus Taipalsaarella Puolustusvoimien ampuma-alueen lähellä on alkanut. Virallinen kuulutus Sarviniemen 80 hehtaarin lomakyläalueen lunastuksesta ilmestyi Puolustusvoimien sivuille perjantaina.

Linkki osoitteeseen:
<http://paikkatiedot.fi/so/11111111/10365897>

Lunastettavia tontteja on kuusitoista. Ne kaikki kuuluvat kahdelle lappeenrantalaissijoittajalle ja heidän yhtiöilleen. Sarviniemessä on lisäksi kymmenen yksityishenkilöiden omistamaa kesämökkiä, mutta niitä Puolustusvoimat ei halua.

Kuva 30: Esimerkki Helsingin Sanomien uutisartikkelista. Lähde: kuvakaappaus Helsingin Sanomien verkkopalvelusta 6.7.2015 (<http://www.hs.fi>).

URI-osoitteista tarjottava RDF-muotoinen tieto mahdollistaa URI-viittauksiin perustuvien tietojen käyttämisen web-sovelluksessa. Web-sovelluksessa voidaan näyttää esimerkiksi ajantasaiset erikieliset nimet pelkän URI-viittauksen avulla. Tätä voidaan hyödyntää mm. erilaisissa tietopalveluissa. Kuvassa 31 on esimerkki Järviwiki-palvelun tietokortista, johon on lisätty ruotsinkielinen nimitieto. Tieto haetaan järveä vastaavan Paikka-kohteen URI-tunnukseen tehdyllä JavaScript-koodiin määritetyllä AJAX-kutsulla. Paikannimet linkitettynä tietona -palvelun hyödyntäminen web-sovelluksissa on yksinkertaista, sillä esimerkiksi palvelun tuottaman JSON-LD-muodon käsittelyssä on mahdollista käyttää JSON-muodossa olevan tiedon käsittelyyn tarkoitettuja JavaScript-työkaluja.

PNR:n aineistosisällön linkittäminen muuhun webin linkitettyyn tietoon mahdollistaa jatkossa monenlaiset kyselyt, joissa eri aineistojen sisältöä voidaan hyödyntää. Hauissa

Suomi Svenska English
Sida Diskussion Visa wikitext Historik

Huvudsida
Vad är Järviwiki?

- Om Järviwiki
- Användarvillkor
- Upphovsrätt
- Integritetspolicy
- Hjälp

Sjörar enligt

- Nimen mukaan
- Avrinningsområde
- Vattenförvaltningsområde

Nuuksion Pitkäjärvi (81.057.1.030)

Nuuksion Pitkäjärvi är en ganska stor insjö i **Suomenlahden rannikkoalue (81)** huvudavrinningsområde. Sjön är belägen i **Nylands landskap**. Den hör till **ELY-centralen i Nylands** verksamhetsområde.

Insjön	Basuppgifter	Administrativa områden
Namn: Nuuksion Pitkäjärvi	Yta: 245,23 ha	Kommun: Espoo
Namn: Noux Långträsk	Djup: 18 m	Landskap: Uudenmaan maakunta
Sjönumm: 81.057.1.030	Volym: 6,48 m	ELY-central: Uudenmaan elinkeino-, ympäristökeskus
Avrinningsområde: Mankinjoen	Volym: 15 893 600 m³	Vattenförvaltningsområde:
valuma-alue (81.057)	Strandlinje: 17,19 km	
Huvudavrinningsområde:	Höjd över havet: 27,3 m	
Suomenlahden rannikkoalue		

Kuva 31: Esimerkki Järviwiki-palvelun tietokortista. Lähde: muokattu kuvakaappaus Järviwiki-verkkopalvelusta 7.7.2015 (<http://www.jarviwiki.fi/>).

voidaan hyödyntää myös spatiaalisia kyselyitä, jolloin paikkojen nimillä voidaan hakea johonkin seutuun liittyviä tietoviittauksia.

PNR-aineiston tarjoaminen linkitettyä tietona avaa sovelluskehittäjille uusia mahdollisuuksia hyödyntää aineistoa kehitystyössä ketterämmin. Tällä hetkellä sovelluksissa hyödynnetään paljon esimerkiksi PNR-aineiston kanssa kilpailevaa GeoNames (2015) -aineistoa, joka on saatavilla mm. API-rajapintojen kautta, mutta myös linkitettyä tietona.

5.5 Ratkaisun arviointi ja jatkokehityssajatuksia

Prototyypipalvelussa on luotu ontologia, jolla PNR-aineiston Paikka- ja Paikannimi-kohteet voidaan esittää RDF-tietona. RDF-tieto tarjotaan tietokohteille annettujen yksilöllisten URI-tunnusten mukaisista osoitteista. RDF-tietosisältö luodaan dynaamisesti MML:n olemassa olevaa WFS-kohdepalvelua hyödyntäen. Lisäksi aineistolle on annettu oma URI-tunnus ja se on jaettu osa-aineistoihin. Aineistot ja osa-aineistot tarjotaan myös RDF-tietona. Kyseisessä RDF-tiedossa kaikki aineiston sisältämät kohteet linkittyvät toisiinsa ja ovat näin saavutettavissa aineisto-URI-tunnuksen kautta. Tässä aliluvussa tuodaan esille palvelun toteutuksessa tehtyjä ratkaisuja, arvioidaan ratkaisujen toimivuutta ja pohditaan PNR-aineiston saavutettavuutta ja käytettävyyttä parantavia kehitysideoita.

Prototyypipalvelu on luotu ketterin menetelmin pala kerrallaan uusien ideoiden myötä. Näin ollen ohjelman rakennetta ei ole suunniteltu alusta loppuun asti. Tämä aiheuttaa joitain ongelmia palvelun ylläpidon, päivittämisen ja laajentamisen osalta. Myöskään testausta ei palvelulle ole suunniteltu, eikä toteutettu. Palvelu toimii kuitenkin hyvänä kokeiluversiona vastaavan tuotantopalvelun kehittämiseksi.

Paikannimet linkitettynä tietona -palvelussa on käytetty ohjelmointiratkaisua, jossa RDFLib-kirjaston avulla graafiin lisätään yksitellen tietoja, joita ei ole WFS-kohdepalvelusta tai XML-skeematiedostoista saatavissa. Ratkaisu ei ole paras mahdollinen. Ainoastaan dynaamisesti skeemoista tai WFS-kohdepalvelun kautta luotava linkitetty tieto kannattaa käsitellä ohjelmallisesti. Muut tiedot (esimerkiksi ”owl:imports”-predikaatilla ilmaistavat suhteet toisiin graafeihin, käytettävät nimiavaruuslyhenteet ja skeematiedostoista puuttuvat tiedot, kuten arvojoukkojen nimet) voidaan kirjoittaa suoraan Turtle-muodossa ja tallentaa tiedostoiksi. Näitä Turtle-dokumentteja voidaan käyttää ohjelmallisesti eräänlaisina valmiiksi täytettyinä pohjina luodessa dynaamista sisältöä. Tällöin sisältötieto on keskitetysti hallittavissa kyseisissä Turtle-dokumenteissa ja ohjelmakoodi pysyy selkeämpänä.

PNR:n Paikat -rajapintatuotteen arvojoukkojen ja luokkien nimiä ei ole käännetty englanniksi alkuperäisessä PNR-dokumentaatioissa tai skeemakuvauksissa. Tätä prototyypipalvelua varten on luotu ruotsin- ja englanninkielisiä käännöksiä arvojoukkojen ja luokkien nimille. Myös ontologiaa varten luoduille assosiaatiosuhteita kuvaaville predikaateille eli PNR-ontologian ”owl:ObjectProperty”- ja ”owl:DatatypeProperty”-luokkien instansseille on annettu erikieliset nimet. Nimiä ei ole tarkistettu, eivätkä ne ole virallisia. Olisi jatkoon kannalta hyvä, jos PNR:n Paikat -rajapintatuotteen kuvausta voisi laajentaa sisältämään kattavammin myös luokkien ja arvojoukkojen erikieliset nimet sekä mahdolliset kuvaukset. Tällöin näitä nimiä ja kuvauksia voitaisiin hyödyntää suoraan myös RDF-tiedossa.

Harvey ym. (2014, s. 5) ehdottavat, että URI-tunnuksissa tulisi näkyä aineiston hierarkkisuus. Tässä palvelussa ei Paikka-kohteiden URI-tunnuksissa ole kuitenkaan käytetty osa-aineistojen mukaisia aineistotunnuksia, vaan jokaisen Paikka-kohteen aineistotunnus on ”11111111”. Tämä siksi, että kohteen kunta- sekä paikkatyyppitieto löytyy joka tapauksessa suorittamalla HTTP-pyyntö kyseiseen URIin. URI-tunnukset ovat näin myös lyhyempiä. Lisäksi lienee mahdollista, että paikkatietokohteen kunta tai paikkatyyppi voi muuttua (esimerkiksi kuntaliitosten yhteydessä tai kunnan saadessa kaupunkistatuksen), vaikka kohde sinänsä ei muutu. Hierarkkinen jako saattaisi näin särkeä, jos pysyväksi tarkoitettua URI-tunnusta ei muutettaisi.

Paikkatiedon tarjoaminen URI-tunnusten perusteella ei riitä, jos tietoa pitää tehokkaasti hakea ja käsitellä. Tiedon hakeminen, käsittely ja yhdistely pelkän linkitysten avulla on hidasta, sillä jokaista haettua kohdetta kohden täytyy tehdä vähintään yksi HTTP-pyyntö. On kuitenkin mahdollista, että halutaan käsitellä tai tutkia esimerkiksi kymmeniä tuhansia kohteita. Ratkaisuna ongelmaan on RDF-tietokannan luominen. RDF-tietokantaa voidaan hyödyntää SPARQL-rajapintapalvelun kautta. SPARQL-palvelu mahdollistaa monipuoliset kyselyt tietoaaineistolle. SPARQL-palvelun tarvitsema RDF-tietokanta olisi mahdollista luoda erikseen esimerkiksi PNR:n tietokannasta tuotetusta GML-muotoisesta sisältötiedostosta tai käyttäen PNR:n WFS-kohdepalvelua kohteiden hakemiseen. Kaikkien kohteiden hakeminen kohdepalvelusta helpottuisi, jos palvelussa olisi käytössä WFS 2.0 -standardin mukainen vastauksen sivutus (engl. response paging, OGC 2010b, s. 29–30). Tällä hetkellä kohdepalvelusta saa korkeintaan 10000 kohdetta yhdellä WFS-kyselyllä.

Aineistolinkityksen toteuttaminen kunnittain- ja paikkatyypeittäin on tehnyt mahdolliseksi aineiston selaamisen web-selaimella linkkien välityksellä. Vastaavanlaista aineiston linkitystä olisi mahdollisesti voinut rakentaa GeoSPARQL-sanaston ”geo:sfWithin”- ”geo:sfContains”-ominaisuuksien avulla. Nyt paikoilla on tieto kunnista, joissa ne sijaitsevat, kunnilla tieto sijaintimaakunnasta ja maakunnilla sijaintiläänistä. Tieto ilmaistaan ”geo:sfWithin”-ominaisuudella. Käänteistä ominaisuutta ”geo:sfContains” ei sen sijaan ole palvelussa hyödynnetty, sillä jotkut kunnat sisältää yli 10000 Paikka-kohdetta. GeoNames.org-palvelussa (2015) asia on ratkaistu luomalla oma ”sisältöresurssi”, jonka URI-tunnuksen mukainen osoite palauttaa listan kohteista, jotka sijaitsevat toisessa paikkatietokohteessa. Tähän sisältöresurssiin viitataan paikkatietokohteesta GeoNames-palvelun oman ontologian ”childrenFeatures”-predikaatilla. Itse paikkatietokohteen URI-tunnuksen palauttavat tiedot pysyvät näin hallittavan kokoisena palauttaen ainoastaan kyseisen viittauksen.

Aineistolinkitys on mahdollistanut myös sen, että hakukoneet voivat indeksoida kaikki eri Paikka- ja Paikannimi-kohteita esittävät web-sivut. Esimerkiksi Googlen sivustonhallintapaneelin kautta voidaan syöttää aineisto-URI, jonka avulla kaikki kohteet päätyvät indeksoitavaksi. Sivustonhallintapaneelin kautta olisi mahdollista syöttää myös sivustokartta eli lista kaikista palvelun sisältämistä web-sivuista. Tämän työn yhteydessä kokeiltun aineisto-URI-tunnuksen syöttämisen jälkeen Googlen indeksointi-nopeus on noussut aina 24000 sivuun päivässä (Googlen sivustohallintapaneelin tilasto 5.10.2015). Tosin tälläkin vauhdilla kaikkien Paikka-kohteita esittelevien sivujen läpikäymiseksi Googlella menee noin 30 päivää. Lisäksi web-sivuja liittyy vielä enemmän Paikannimi-kohteisiin kuin Paikka-kohteisiin.

Schema.org -sanaston mukainen Paikka-kohteiden RDF-muotoinen esitys tarjotaan liitettynä HTML-sivun lähdekoodin ”script”-elementin sisään. Lisäksi se on saatavilla Paikka-kohteen URI-tunnuksen mukaisesta osoitteesta lisäämällä osoitteeseen ”.jsonld”-pääte. Schema.org-sanaston mukaisen RDF-kuvauksen tarjoaminen onnistuisi mahdollisesti myös Content Negotiation -tekniikalla. Ongelma on se, että Content Negotiation -tekniikan hyödyntämä JSON-LD-muodon tietoformaattitunnus ”application/json+ld” on varattu PNR-ontologian mukaiselle JSON-LD-muotoiselle RDF-kuvaukselle. Schema.org- ja PNR-ontologian muotoisia kuvauksia ei kannattane sisällyttää samaan tarjottavaan graafiin, koska tällöin toisteinen tieto lisääntyisi ja OWL DL -päätelyssä syntyisi mahdollisesti ristiriitoja.

6 Yhteenveto

Tämän työn tutkimuskysymyksiin on vastattu toteuttamalla prototyyppipalvelu, joka tarjoaa paikkatietoaineiston kohteet linkitettynä tietona. Prototyyppipalvelua voidaan kuvata onnistuneeksi. Se tarjoaa PNR-aineiston RDF-tietona dynaamisesti hyödyntäen olemassa olevaa MML:n WFS-kohdepalvelua. Kaikkea palvelussa tarvittavaa tietoa ei luoda reaaliaikaisesti. PNR-aineiston tyyppiluokitusta, XML-skeemaa sekä kuntia, maakuntia ja läänejä vastaavia Paikka-kohteita hyödynnetään luvussa 5.3.3 esitellyissä kerta-ajoprosesseissa. Kerta-ajoprosessit vaikuttavat mm. ontologiasa käytettävien termien nimiin, paikkatyyppihierarkiaan ja GeoSPARQL-sanaston topologiasuhteella ”geo:sfWithin” ilmaistuihin Paikka-kohteiden kunta-, maakunta- ja läänisuhteisiin. Kerta-ajoprosesseissa konvertoidut tiedot ovat kaikki luonteeltaan melko pysyviä, jolloin odotettuja muutoksia on erittäin harvoin. Sen sijaan yksittäisten kohteiden tietosisältö haetaan WFS-kohdepalvelusta ja muunnetaan reaaliaikaisesti linkitetyksi tiedoksi. Tällöin kohdepalvelussa tehtävät muutokset välittyvät ajantasaisesti myös Paikannimet linkitettynä tietona -palveluun. Reaaliaikaiset prosessit hyödyntävät kerta-ajoprosesseissa luotua tietoa, jolloin reaaliaikaiset prosessit ovat huomattavasti kevyempiä ja nopeampia.

PNR-aineiston paikkatietokohteet ovat luodussa palvelussa saatavilla yksilöllisten URI-tunnusten perusteella. Yksilöllisten URI-tunnusten antaminen paikkatietokohteille onnistuu JHS 193 -suosituksen mukaisesti antamalla aineistolle aineistotunnus ja hyödyntämällä paikkatietokohteiden kohdetunnuksia URI-tunnuksen paikallisena osana. URI-tunnusten toimiminen web-osoitteina onnistuu uudelleenohjaustekniikan ja web-palvelimen ominaisuuksien avulla. Aineiston kaikki paikkatietokohteet ja niiden URI-tunnukset ovat saatavilla aineistolinkityksen avulla. Paikkatietokohteet on linkitetty toisiinsa hyödyntäen VoID-sanastoa ja luomalla ainestoa sekä osa-aineistoja kuvaavat resurssit. Aineiston kaikista resursseista on siis pääsy mihin tahansa toiseen resurssiin linkkien kautta. Näin ollen linkitetyn tiedon neljäs periaate (kts. luku 3.1) täyttyy aineiston sisäisesti. Perusteltua olisi lisäksi tarjota aineisto SPARQL-palvelun kautta. SPARQL-palvelu tarjoaa monipuoliset mahdollisuudet hakea aineistosta tietoa ja yhdistellä aineiston tietoa toisiin aineistoihin. SAPRQL-palvelu on yksinkertainen perustaa tässä työssä luodun ontologian ja muunnosprosessin avulla.

PNR-tietomallin ontologisointi on toteutettu OWL-ontologiakieltä hyödyntäen. Ontologisoinnissa on käytetty Tschirnerin ym. (2011, s. 79) tutkimuksessa kehitettyjä sääntöjä. Ontologiasta on tehty hierarkkinen käyttämällä hyväksi PNR-tietomallin paikkatyyppijakoa. Tämä helpottaa aineiston sisällön jäsentämistä. Ontologian luomisessa on seurattu linkitetyn tiedon periaatteita, jotta aineistolle saavutetaan mahdollisimman paljon lisäarvoa. Ontologia hyödyntää GeoSPARQL-sanastoa niin, että paikkatietokohteiden käsittely mahdollistuu GeoSPARQL-standardin mukaisesti.




Työn yhtenä tavoitteena oli selvittää, miten paikkatietokohteet mahdollisesti näkyvät webin hakukoneiden hakutuloksissa. Tätä on testattu Googlen hakupalvelulla. Syöttämällä Googleen hakusanat ”paikat heinola” saadaan ensimmäisellä sivulla

kolme Paikannimet linkitettynä tietona -palvelun osa-aineistoa näkyviin (kts. kuva 32). Näkyvyys hakukonetuloksissa kiistämättä lisää PNR-aineiston saavutettavuutta, mutta näkymisen hyödyntämismahdollisuuksien selvittäminen vaatii jatkotutkimuksia. HTML-sivuille lisätty schema.org-sanaston mukainen RDF-tieto mahdollistaa kohteiden koneellisen tulkinnan. Googlen sivustohallintapaneelistä voidaan todeta, että Google tunnistaa ”schema:Place”-tyypin kohteet Paikka-kohteita esittävästä HTML-sivusta. Se mahdollistaa esimerkiksi kohteista esitetyn rakenteisen tiedon automaattisen näyttämisen kartalla ja automaattisen sijaintikuntatiedon esittämisen ”schema:containedIn”-predikaatin perusteella.

Palvelun tarjoama aineisto saa tällä hetkellä Berners-Leen (2006) luoman luokitusjärjestelmän mukaan neljä tähteä. Viidennen tähden vaatimus tiedon linkittymisestä ulkoisiin aineistoihin ei toistaiseksi toteudu. Tällainen mahdollisuus olisi esimerkiksi paikannimien linkittäminen MML:n Maastotietokannan vastaaviin kohteisiin. Sen sijaan Hyvönen ym. (2014) ehdottaman laajennoksen mukaisen kuudennen tähden saamiseksi vaadittava skeema on aineistolle räätälöidyn ontologian muodossa jo olemassa.

Paikkatietoaineiston saavutettavuuteen ja etenkin käytettävyyteen vaikuttavat aineistosta saatavilla olevat metatiedot. Paikkatietoaineiston metatietoja on tarjottu prototyypissä luodun VoID-sanaston ”void:Dataset”-luokan mukaisen aineistoresurssin ”<http://paikkatiedot.fi/so/11111111/>” RDF-muotoisessa kuvauksessa. Kuvauksessa hyödynnetään Dublin Core -sanastoa, joka on tarkoitettu aineistojen metatiedon ilmaisemiseksi. Metatietojen merkitystä ei voi liikaa korostaa. Se, onko aineisto saatavilla WFS-kohdepalvelun kautta tai linkitettynä tietona, ei kerro aineiston luotettavuudesta, tarkkuudesta tai sen soveltuvuudesta erilaisiin käyttötarkoituksiin. Molemmista tapauksista avainasemassa ovat aineistosta saatavilla olevat metatiedot.

Linkitetyn tiedon tekniikat tasoittavat tietä paikkatietoaineistojen monipuoliselle hyödyntämiselle ja yhdistämiselle muihin tietoaineistoihin. Tässä työssä luotu prototyypipalvelu hyödyntää paikkatiedon yksilöllisiä URI-tunnuksia aineiston tarjoamiseksi. Web-ympäristössä URI-tunnukset mahdollistavat yksinkertaisen tiedon saannin sekä yksikäsitteisen paikkatietokohteiden linkittämisen tunnusten perusteella. Koneluettavan RDF-tiedon tarjoaminen URI-tunnusten perusteella mahdollistaa tiedon hyödyntämisen webin linkitetyn tiedon käsittelyyn tarkoitettujen työkalujen ja sovellusten avulla.

Google   

[Verkkohaku](#) [Kartat](#) [Kuvahaku](#) [Videot](#) [Lisää ▾](#) [Hakutyökalut](#)

Noin 321 000 tulosta (0,32 sekuntia)

Uimarannat ja -paikat - Heinola
[www.heinola.fi/uimarannat-ja-paikat ▾](http://www.heinola.fi/uimarannat-ja-paikat)
 Heinolan sijainnin ansiosta kaupungissa on useita uimarantoja ja -paikkoja. EU -vesidirektiivin mukaisesti valvottuja uimarantoja on seitsemän: Kylpylän ...

Esiopetuspaikat - Heinola
[www.heinola.fi/esiopetuspaikat ▾](http://www.heinola.fi/esiopetuspaikat)
 Päiväkodeissa tai päiväkotien kouluilla toimivissa ryhmissä lukuunottamatta Myllyojan ryhmäperhepäiväkotia tarjotaan maksutonta esiopetusta lapsille vuotta ...

Päijät-Hämeen TET-paikkoja - Peda.net
[peda.net/veraja/tori/paijat-hame/kaikki ▾](http://peda.net/veraja/tori/paijat-hame/kaikki)
 Päijät-Hämeen TET-paikkoja. Logo ... Elämänkaaritalo, Heinola - Erstan koulun keittiö, Nastola ... Heinolan kaupunki ATK -palvelut - Heinolan kaupunki/ ...

Paikat: Heinola, Metsäalue
[inspire-hy.fgi.fi/doc/11111111:111325/ ▾](http://inspire-hy.fgi.fi/doc/11111111:111325/)
<http://paikkatiedot.fi/so/11111111:111325/> · Paikat: Heinola, Metsäalue. Tietoaineiston kuvaus. in dataset · <http://paikkatiedot.fi/so/11111111:111/>. Sisältää ...

Paikat: Heinola, Niemi - FGI
[inspire-hy.fgi.fi/doc/11111111:111345/ ▾](http://inspire-hy.fgi.fi/doc/11111111:111345/)
 RDF-tietoaineisto. <http://paikkatiedot.fi/so/11111111:111345/> · Paikat: Heinola, Niemi. Tietoaineiston kuvaus. in dataset · <http://paikkatiedot.fi/so/11111111:111/>.

Paikat, Heinola
[inspire-hy.fgi.fi/doc/11111111:111/ ▾](http://inspire-hy.fgi.fi/doc/11111111:111/)
 Paikat: Heinola, Allas, <http://paikkatiedot.fi/so/11111111:111520/>. Paikat: Heinola, Autotie, <http://paikkatiedot.fi/so/11111111:111100/>. Paikat: Heinola, Erämaa- ...

www.heinola.net - Piukat paikat
[www.heinola.net/pages/arkisto/piukat-paikat.php ▾](http://www.heinola.net/pages/arkisto/piukat-paikat.php)
 Ensi-ilta 7.6.2001. Teksti: Peter Stone Sävellys: Jule Styne Laulujen sanat: Bob Merrill Suomennos: Jussi Tuominen Musikaalin ohjaus: Markus Packalén

Kuva 32: Google-hakutuloksia hakusanoilla ”paikat heinola”. Lähde: kuvakaappaus Google-verkkopalvelusta 7.10.2015 (<http://www.google.com>).

Viitteet

- Battle, Robert ja Dave Kolas (2012). "Enabling the geospatial semantic web with parliament and GeoSPARQL". *Semantic Web* 3.4. s. 355–370. IOS Press. URL: http://www.semantic-web-journal.net/sites/default/files/swj176_2.pdf.
- Beckett, David, Tim Berners-Lee, Eric Prud'hommeaux ja Gavin Carothers (2015). "RDF 1.1 Turtle - Terse RDF Triple Language. W3C Recommendation 25 February 2014". Toim. Gavin Carothers Eric Prud'hommeaux. W3C:n standardi. URL: <http://www.w3.org/2015/spatial/charter>.
- Berners-Lee, T., R. Fielding ja L. Masinter (2005). "RFC 3986: Uniform Resource Identifier(URI): Generic Syntax". STD 66, RFC 3986. Internet Standard, IETF, 61 s. URL: <http://www.rfc-editor.org/std/std66.txt>.
- Berners-Lee, Tim (2006). "Linked Data". Verkkosivu. URL: <http://www.w3.org/DesignIssues/LinkedData.html> (viitattu 10.08.2015).
- Butler, Howard, Martin Daly, Allan Doyle, Sean Gillies, Tim Schaub ja Christopher Schmidt (2008). "The GeoJSON format specification". Tekninen raportti, 67 s. URL: <http://geojson.org/geojson-spec.html>.
- CIDOC, Center for Intercultural Documentation (2015). "CIDOC Conceptual Reference Model". Verkkosivu. URL: <http://www.cidoc-crm.org> (viitattu 21.09.2015).
- Cox, Simon (2013). "An explicit OWL representation of ISO/OGC Observations and Measurements." *6th International Workshop on Semantic Sensor Networks. In conjunction with the 12th International Semantic Web Conference (ISWC)*. 18 s. URL: <http://ceur-ws.org/Vol-1063/paper1.pdf>.
- DCMI Usage Board (2015). "DCMI Metadata Terms". Dublin Core Metadata Initiative (DCMI) suositus. Verkkosivu. URL: <http://dublincore.org/documents/2012/06/14/dcmi-terms/> (viitattu 26.08.2015).
- ESRI, Environmental Systems Research Institute (1998). "Shapefile technical description". *An ESRI White Paper*. Tekninen raportti, 34 s. URL: <https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>.
- Euroopan komissio (2015). "Implementation of Identifiers using URIs in INSPIRE". Verkkosivu. URL: <http://inspire.ec.europa.eu/ids> (viitattu 15.09.2015).
- EY, Euroopan yhteisö (2007). "Euroopan parlamentin ja neuvoston direktiivi 2007/2/EY, annettu 14 päivänä maaliskuuta 2007, Euroopan yhteisön paikkatietoinfrastruktuurin (INSPIRE) perustamisesta". *Euroopan unionin virallinen lehti*. L 108 25.4.2007, s. 1–14. URL: <http://eur-lex.europa.eu/legal-content/FI/TXT/PDF/?uri=CELEX:32007L0002&from=FI>.

- GeoNames (2015). "GeoNames". Verkkosivu. URL: <http://www.geonames.org> (viitattu 02.09.2015).
- Google (2015). "About schema.org". Verkkosivu. URL: <https://developers.google.com/structured-data/schema-org> (viitattu 01.10.2015).
- GSDI, Global Spatial Data Infrastructure Association (2004). "Developing Spatial Data Infrastructures: The SDI Cookbook". Versio 2.0. Toim. Douglas D. Nebert. URL: http://gsdiassociation.org/images/publications/cookbooks/SDI_Cookbook_GSDI_2004_ver2.pdf.
- Guldborg, Erik (2015). "Vi prøver oss på linked open data". Kartverket Norge. Verkkosivu. URL: <http://labs.kartverket.no/vi-prover-oss-pa-linked-open-data/> (viitattu 20.07.2015).
- Harvey, Francis ym. (2014). "Little steps towards big goals. Using linked data to develop next generation spatial data infrastructures (aka SDI 3.0)". *Proceedings of the AGILE'2014 International Conference on Geographic Information Science, Castellón, June, 3-6, 2014*. AGILE Digital Editions, 6 s. URL: http://www.agile-online.org/Conference_Paper/cds/agile_2014/agile2014_110.pdf.
- Heath, Tom ja Christian Bizer (2011). "Linked data: Evolving the web into a global data space". *Synthesis lectures on the semantic web: theory and technology* 1.1. 136 s. Morgan & Claypool Publishers. URL: <http://linkeddatabook.com/editions/1.0/>.
- Hietanen, Eero (2015). "Paikannimet linkitettynä tietona". Maanmittauslaitoksen paikkatietokeskus FGI. Verkkosivu. URL: <http://paikkatiedot.fi/so/11111111/> (viitattu 18.09.2015).
- Hyvönen, Eero, Jouni Tuominen, Miika Alonen ja Eetu Mäkelä (2014). "Linked Data Finland: A 7-star model and platform for publishing and re-using linked datasets". *The Semantic Web: ESWC 2014 Satellite Events*. s. 226–230. Springer. URL: <http://www.seco.tkk.fi/publications/2014/hyvonon-et-al-ldf-2014.pdf>.
- INSPIRE Drafting Team: Data Specifications (2013). "D2.5: Generic Conceptual Model, Version 3.4rc3". D2.5_v3.4rc3. Tekninen raportti 157s. URL: http://inspire.ec.europa.eu/documents/Data_Specifications/D2.5_v3.4rc3.pdf.
- ISO (2005). "Geographic information – Rules for application schema". ISO 19109:2005. Standardi, 79s. Suomen Standardisoimisliitto SFS, Helsinki.
- Jones, Jim, Werner Kuhn, Carsten Kessler ja Simon Scheider (2014). "Making the web of data available via web feature services". *Connecting a Digital Europe Through Location and Place*. s. 341–361. Springer International Publishing. URL: <http://carsten.io/jones-kuhn-kessler-scheider-agile2014.pdf>.

- JUHTA, Julkisen hallinnon tietohallinnon neuvottelukunta (2015). ”JHS 193 Paikkatiedon yksilöivät tunnukset”. JHS-suositus. URL: <http://www.jhs-suositukset.fi/suomi/jhs193>.
- Koistinen, Kai (2011). ”Paikkatiedon tietotuoteskeemojen ontologisointi tiedonhaun tueksi”. Diplomityö, 57 s. Aalto-yliopisto, Insinööritieteiden korkeakoulu. URL: http://maa.aalto.fi/fi/research/gma/geoinformatics_and_cartography/masters/.
- Leskinen, Teemu (2015). ”A Data Repository for Named Places and Their Standardised Names Integrated With the Production of National Map Series”. *7th International Cartographic Conference. 16th General Assembly August 23-28, 2015 Rio de Janeiro, Brazil. Maps Connecting the World*. 15 s. URL: http://icaci.org/files/documents/ICC_proceedings/ICC2015/papers/21/516.html.
- Lindroos, Robin (2008). ”Paikkatiedon ontologiapalvelu”. Diplomityö, 74 s. Teknillinen Korkeakoulu. URL: <http://www.seco.tkk.fi/publications/2008/onkipaikka-diplomityo-2008.pdf>.
- Linked Data Finland (2015). ”Finnish Geographic Names”. Verkkosivu. URL: <http://www.ldf.fi/dataset/pnr/index.html> (viitattu 23.07.2015).
- MML, Maanmittauslaitos (2015). ”Aineistot ja tuotteet”. Maanmittauslaitos. Verkkosivu. URL: <http://www.maanmittauslaitos.fi/aineistot-palvelut/rajapintapalvelut/nimiston-kyselypalvelu-wfs/kayttoonotto/aineistot-tuotteet> (viitattu 18.09.2015).
- Norton, Barry ym. (2012). ”NeoGeo Vocabulary Specification - Madrid Edition”. Toim. Juan Martín Sala ja Andreas Harth. Verkkosivu. URL: <http://geovocab.org/doc/neogeo-20120205/> (viitattu 27.08.2015).
- OGC, Open Geospatial Consortium (2005). ”OpenGIS® Implementation Specification for Geographic information - Simple feature access - Part 1:Common architecture”. Versio 1.1.0. Toim. Keith Ryden. OGC 05-126 ja ISO 19125. Standardi, 51 s. URL: http://portal.opengeospatial.org/files/?artifact_id=13227.
- OGC, Open Geospatial Consortium (2006). ”OpenGIS® Web Map Server Implementation Specification”. Versio 1.3.0. Toim. Jeff de la Beaujardiere. OGC® 06-042. Standardi, 85 s. URL: http://portal.opengeospatial.org/files/?artifact_id=14416.
- OGC, Open Geospatial Consortium (2010a). ”OpenGIS Filter Encoding 2.0 Encoding Standard”. Toim. Peter Vretanos. OGC 09-026r1 ja ISO/DIS 19143. Standardi, 90 s. URL: http://portal.opengeospatial.org/files/?artifact_id=39968.
- OGC, Open Geospatial Consortium (2010b). ”OpenGIS Web Feature Service 2.0 Interface Standard”. Toim. Panagiotis A. Vretanos. OGC 09-025r1 ja ISO/DIS 19142. Standardi, 253 s. URL: http://portal.opengeospatial.org/files/?artifact_id=39967.

- OGC, Open Geospatial Consortium (2012a). "OGC® Geography Markup Language (GML) — Extended schemas and encoding rules". Versio 3.3. Toim. Clemens Portele. OGC 10-129r1. Standardi, 91 s. URL: <http://www.opengis.net/spec/GML/3.3>.
- OGC, Open Geospatial Consortium (2012b). "OGC GeoSPARQL - A geographic query language for RDF data". Versio 1.0. *OGC Implementation Standard*. Toim. Matthew Perry ja John Herring. OGC 11-052r4. Standardi, 75 s. URL: <http://www.opengis.net/doc/IS/geosparql/1.0>.
- OGC, Open Geospatial Consortium (2014). "OGC® GeoPackage Encoding Standard". Versio 1.0. *OGC® Encoding Standard*. Toim. Paul Daisey. OGC 12-128r10. Standardi, 153 s. URL: <http://www.opengis.net/doc/IS/geopackage/1.0>.
- Schade, Sven ja Simon Cox (2010). "Linked Data in SDI or How GML is not about Trees". *Proceedings of the 13th AGILE International Conference on Geographic Information Science-Geospatial Thinking*. 10 s. URL: http://www.agile-online.org/Conference_Paper/CDs/agile_2010/ShortPapers_PDF/73_DOC.pdf.
- Schade, Sven, Carlos Granell ja Laura Díaz (2010). "Augmenting SDI with linked data". *Workshop On Linked Spatiotemporal Data, in conjunction with the 6th International Conference on Geographic Information Science (GIScience 2010). Zurich, 14th September*. 12 s. URL: <http://ceur-ws.org/Vol-691/paper3.pdf>.
- schema.org (2015). "schema.org". Verkkosivu. URL: <http://schema.org/Thing> (viitattu 26.08.2015).
- Sporny, Manu, Dave Longley, Gregg Kellogg, Markus Lanthaler ja Niklas Lindström (2014). "JSON-LD 1.0 - A JSON-based Serialization for Linked Data. W3C Recommendation 16 January 2014". Toim. Manu Sporny, Gregg Kellogg ja Markus Lanthaler. W3C:n standardi. URL: <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
- Tschirner, Sven, Ansgar Scherp ja Steffen Staab (2011). "Semantic access to INSPIRE". *Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web* 798. s. 75–87. CEUR-WS.org. URL: <http://ceur-ws.org/Vol-798/>.
- W3C, World Wide Web Consortium (2006a). "Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation 16 August 2006, edited in place 29 September 2006". Toim. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau ja John Cowan. W3C:n standardi. URL: <http://www.w3.org/TR/2006/REC-xml11-20060816/>.
- W3C, World Wide Web Consortium (2006b). "Semantic Web Interest Group: Basic Geo (WGS84 lat/long) Vocabulary". Toim. Dan Brickley. Verkkosivu. URL: <http://www.w3.org/2003/01/geo/> (viitattu 27.08.2015).

- W3C, World Wide Web Consortium (2007). "W3C Geospatial Vocabulary. W3C Incubator Group Report 23 October 2007". Toim. Joshua Lieberman, Raj Singh ja Chris Goad. Tekninen raportti. URL: <http://www.w3.org/2005/Incubator/geo/XGR-geo-20071023/> (viitattu 27.08.2015).
- W3C, World Wide Web Consortium (2008). "Cool URIs for the Semantic Web. W3C Interest Group Note 03 December 2008". Toim. Leo Sauermann ja Richard Cyganiak. Verkkosivu. URL: <http://www.w3.org/TR/2008/NOTE-cooluris-20081203/> (viitattu 28.07.2015).
- W3C, World Wide Web Consortium (2009). "SKOS Simple Knowledge Organization System Reference. W3C Recommendation 18 August 2009". Toim. Alistair Miles ja Sean Bechhofer. W3C:n standardi. URL: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.
- W3C, World Wide Web Consortium (2012a). "OWL 2 Web Ontology Language Direct Semantics (Second Edition). W3C Recommendation 11 December 2012". Toim. Boris Motik, Peter F. Patel-Schneider ja Bernardo Cuenca Grau. W3C:n standardi. URL: <http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/>.
- W3C, World Wide Web Consortium (2012b). "OWL 2 Web Ontology Language Primer (Second Edition). W3C Recommendation 11 December 2012". Toim. Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider ja Sebastian Rudolph. Verkkosivu. URL: <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/> (viitattu 26.08.2015).
- W3C, World Wide Web Consortium (2012c). "OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition). W3C Recommendation 11 December 2012". Toim. Michael Schneider. W3C:n standardi. URL: <http://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/>.
- W3C, World Wide Web Consortium (2012d). "W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures. W3C Recommendation 5 April 2012". Toim. Shudi Gao, C. M. Sperberg-McQueen, Henry S. Thompson, Noah Mendelsohn, David Beech ja Murray Maloney. W3C:n standardi. URL: <http://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/>.
- W3C, World Wide Web Consortium (2013a). "SPARQL 1.1 Overview. W3C Recommendation 21 March 2013". Toim. Carlos Buil Aranda ym. W3C:n standardi. URL: <http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>.
- W3C, World Wide Web Consortium (2013b). "W3C Semantic Web Activity". Toim. Sandro Hawke, Iwan Herman, Phil Archer ja Eric Prud'hommeaux. Verkkosivu. URL: <http://www.w3.org/2001/sw/> (viitattu 04.08.2015).
- W3C, World Wide Web Consortium (2014a). "RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014". Toim. Richard Cyganiak, David Wood, Markus Lanthaler, Graham Klyne, Jeremy J. Carroll ja Brian

- McBride. W3C:n standardi. URL: <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
- W3C, World Wide Web Consortium (2014b). "RDF 1.1 Primer. W3C Working Group Note 24 June 2014". Toim. Guus Schreiber, Yves Raimond, Frank Manola, Eric Miller ja Brian McBride. Verkkosivu. URL: <http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/> (viitattu 24.08.2015).
- W3C, World Wide Web Consortium (2014c). "RDF 1.1 XML Syntax. W3C Recommendation 25 February 2014". Toim. Fabien Gandon, Guus Schreiber ja Dave Beckett. W3C:n standardi. URL: <http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>.
- W3C, World Wide Web Consortium (2015a). "Semantic Web". Verkkosivu. URL: <http://www.w3.org/standards/semanticweb/> (viitattu 04.08.2015).
- W3C, World Wide Web Consortium (2015b). "Spatial Data on the Web Working Group Charter". Versio 1.10. Toim. Phil Archer. Verkkosivu. URL: <http://www.w3.org/2015/spatial/charter> (viitattu 23.07.2015).
- YK, Yhdistyneet Kansakunnat (1992). "AGENDA 21". *United Nations Conference on Environment & Development Rio de Janeiro, Brazil, 3 to 14 June 1992*. Toimintaohjelma, 351 s. URL: <https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf>.